A Hybrid Analysis of the State of Automated Journalism in Canada: Current Impact and Future Implications for Journalists and Newsrooms

Brigitte Tousignant

A Thesis

in the Department

of

Journalism

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Journalism Studies) at

Concordia University

Montreal, Quebec, Canada

AUGUST 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Brigitte Tousignant

Entitled: A Hybrid Analysis of the State of Automated Journalism in Canada: Current Impact and Future Implications for Journalists and Newsrooms

and submitted in partial fulfillment of the requirements for the degree of

Masters of Arts (Digital Innovation in Journalism Studies)

complies with the regulations of the University and meets the accepted standards with respect to the originality and quality.

Signed by the final examining committee:

Dr. Andrea Hunter _____ Chair

Dr. Elyse Amend _____ Supervisor

Dr. David Secko _____ Examiner

Approved by: _____
Dr. David Secko
Chair of the Department or Graduate Program Director

_____
Dr. André Roy
Dean of Faculty

Date: _____

ABSTRACT

A Hybrid Analysis of the State of Automated Journalism in Canada:

Current Impact and Future Implications for Journalists and Newsrooms

Brigitte Tousignant

In recent years, certain media organizations around the world have begun adopting automated journalism to enhance their newsrooms' productivity and reporting capacity. These tools, which can be fully- or semi-autonomous expert systems, are being implemented for a multitude of reasons, such as to ease journalists' workflow, to expand news coverage, to uncover complex investigations and to cut costs. While some have argued automation can help increase newsroom efficiency and output, questions regarding authorship and transparency, and the implications for journalists, have come into question. Despite automated journalism's growing presence, little is known about its integration in and influence on the Canadian news media landscape.

This thesis reports on the results of qualitative interviews with nine journalists and news media professionals and aims to examine how automated journalism techniques and tools are being integrated, and the effects they have on journalists, their practices, methods, and what is being demanded of them. The purpose of this thesis is to fill a gap in the literature and in our knowledge about automated journalism's role in Canada. Its findings conclude that, although Canada is behind in the adoption of automated journalism, there is an overwhelming consensus that the ethical guidelines of the Canadian news media industry need to be revised in order to better frame the use of automated technologies.

Acknowledgements

I wish to express my deepest gratitude to my thesis supervisor, Dr. Elyse Amend, for her invaluable guidance, assistance and encouragement throughout my studies. Her dedication toward helping me achieve my research goals throughout this process has tremendously contributed to my academic, professional and personal growth. Although the coronavirus drastically altered the last few months of my research and the 2020 academic year, it is important that I pay special regards to Dr. Amend's patience and unwavering devotedness toward her students despite the many challenges presented this year.

I would also like to thank my thesis committee members, Dr. David Secko and Dr. Andrea Hunter, for their enthusiasm and insight for the duration of my time as a Master's student in Concordia University's Journalism Department. In addition, I would like to thank professor emeritus of the journalism department, Dr. Mike Gasher, who willingly took some time out of his retirement to read my thesis, further contributing to my academic growth.

I am deeply indebted to my friends—those whom I met in the Master's program and those who have been stuck with me for years—for their attentiveness and ability to make me laugh even when the going gets tough. Thank you for always keeping me sane.

Finally, I must express my very profound gratitude to my parents, Renée and Maurice, my brother Bruno, and my twin brother, Philly, for their invaluable love, support, and endless homecooked meals.

Table of Contents

(Continued on next page)

List of Figures

List of Tables

Introduction and Purpose of Thesis

Today, many industries are shifting their practices to carve out spaces for what some have claimed to be the era of "algorithmic revolution in knowledge production," and many sectors of the journalism industry have followed suit (Anderson, 2012). Automation and algorithms have attained a level of sophistication where they now have the ability to permeate the entire news production pipeline, whether generating articles out of structured data, taking over menial tasks to free up journalists to pursue other work, creating interactive media such as news bots, or uncovering complex investigative stories with data-mining methods and machine-learning (Diakopoulos, 2019, p. 3; Carlson, 2015; Broussard, 2015, Graefe, 2016; Dörr, 2016). This thesis explores the adoption of algorithms and automation in Canadian newsrooms and aims to assess the state of automated journalism in Canada, particularly with regards to its practitioners, the implications for journalists, and the editorial guidelines or newsroom policies surrounding this technology.

Known as automated journalism, and sometimes referred to as "robo-journalism" or artificial intelligence-enhanced journalism, this emerging field of newswork stems from artificial intelligence (AI). It derives its skills from computer programs, known as expert systems, which can perform highly sophisticated tasks with little-to-no human input or interaction—aside from their original coding (Broussard, 2015, Lokot & Diakopoulos, 2015, Diakopoulos, 2019). Many of the leading media organizations in the United States, like the *Washington Post, LA Times, Associated Press* and ProPublica are now using automated journalism to generate articles at scale, with quick turnaround times and low-marginal costs (Carlson, 2015; Montal & Reich, 2016; Lokot & Diakopoulos, 2016; Graefe, 2016).

Despite the growing number of news outlets that are incorporating this technology, and a steady rise in companies developing or funding software solutions for automated news (Graefe, 2016), there exists a number of conflicting views over the use of algorithmically-generated content in the field, including the degree to which journalists fear displacement (Carlson, 2015; Graefe, 2016), challenges in accountability (Diakopoulos, 2016; O'Neil, 2016, p. 12; Dörr & Hollnbuchner, 2017), questions of authorship and transparency (Montal & Reich, 2016; Broussard, 2018, p. 194), as well as nuances regarding the accuracy and subsequent consequences it may have for democracy and society at large (Diakopoulos, 2016; Wolley, et al., 2016; Noble, 2018, p. 4).

Moreover, while the literature about automated journalism has increased in past years (Montal & Reich, 2016), little is known about the state of automated journalism in Canadian newsrooms, particularly with regards to its practitioners, implications for Canadian journalists and the editorial guidelines or newsroom policies surrounding these technologies. To date, there have been no Canada-focused research contributions in this area, despite Canada cementing its position as a global leader in AI research and development. For instance, in 2017, the federal government ratified the importance of AI in the Canadian economy by funding a $125 million Pan-Canadian AI Strategy for research, of which $40 million was allocated to AI industry members and businesses in Montreal, Quebec. That same year, the Government of Quebec issued $100 million to its AI ecosystem.

Yet with regard to the use of artificial intelligence-enhanced journalism (Diakopoulos, 2019), the bulk of the research has to date focused on US and European newsrooms (Dörr, 2016; Montal & Reich, 2016; Graefe, 2016), despite the recent uptake of automated journalism by Canadian news organizations. For example, the *Canadian Press* has fully automated their junior

hockey games recaps, generating upwards of 200 sports scores per week, covering hockey games in all of the provinces. The statistics for every game are plugged into a natural language generation software, a process that transforms structured data into readable content, which is then automatically published online. It is done without any human input apart from the initial programming of the software. At the *Globe and Mail,* they developed an algorithmically-driven tool that allows their reporters to build their own graphic and visualization charts in under 60 seconds to complement their news articles—a process which they claim has "democratized" this type of newswork by eliminating the need to go through several people to obtain data visualizations (Journalist 9, personal communication, March 4 2020). Still, no known critical research has investigated who in Canadian media are implementing AI, and the effects this has on the roles of Canadian journalists, their practices, methods, and tools. These are the questions that drive this thesis and the incentives in pursuing this research in order to fully comprehend where the country's news media industry stands regarding the integration of this technology.

By conducting qualitative interview research, this thesis seeks to contribute to the knowledge of journalism research and practice through the lived experiences and detailed views of Canadian journalists and other news media professionals working for French- and English-language news media outlets. Newsrooms and journalists considered for participation in this research were those who use automation or semi-autonomous tools to enhance their reporting, including, but not limited to, content optimization; data-mining and scraping; news bots to interact with audiences online; content generation; and those who complement data journalism with enhanced automation methods.

Four main broad research questions will guide this thesis:

RQ1. Who is exercising automated journalism in Canada?

RQ2. What types of news content are they producing and to what capacity?

RQ3. What are the realities of the journalists working alongside automated journalism?

RQ4. What reasons motivate the adoption of AI in newsrooms?

The thesis is structured as follows:

Chapter One provides an in-depth review of the existing literature, beginning with an overview of AI's earliest adoptions and the mathematicians and philosophers who started positing whether machines, namely digital computers, could demonstrate cognitive intelligence equal to, or surpassing, a human's capacity (Turing, 1950; Dreyfus, 1993; Haugeland, 1989). Early academic research into the subject has long argued that a machine's capacity will never match the level of sophistication of the human's, because computers will always lack the ability to be creative and critical (Dreyfus, 1993; Haugeland, 1989). Still, it is important to juxtapose past research with modern-day strides in artificial intelligence, as recent developments in automated technologies have some journalists fearing some degree of displacement (Carlson, 2015; Graefe, 2016; Stray, 2016).

This section then dives into the many different applications of automated journalism throughout the news production chain, and presents a reflection of its influence and changes on journalistic values in the development, integration and uses of these technologies. This chapter then discusses the impact of the widespread adoption of algorithms, and their effects on society at large, and the urgent need to reassess transparency and accountability guidelines (Broussard, 2016; Eubanks, 2018; Woolley, Boyd & Broussard, 2016). Finally, it sheds light on the existing body of knowledge with regards to the role these technologies play in enhancing, or undermining, the journalistic business model.

Chapter Two begins by first identifying the important theoretical ideas that inform this thesis and explores the mutual constitutions of journalism and technology through the emerging field of study known as Human-Machine-Communication (Guzman, 2018). As Diakopoulos (2019) puts it, "the history of journalism is one of adaptation as news technologies—telephony, photography, reproduction, and computerization—changed the nature of roles, tasks and workflows" (p. 5). By highlighting Lewis and Westlund's (2015) call for a more holistic approach, this thesis studies the diverse relationships between journalism and technology, which encompasses journalists, technologists and technological actants. It then provides a definition of automated journalism as an area of computational journalism, and further considers the potential roles and limitations of automation within Canadian journalism, as an industry that is constantly competing to turn up captivating and original story ideas in today's digital age, while also expanding its coverage and dealing with the reality of dwindling resources.

Chapter Three explains this research's selected methodology and provides a detailed overview of its findings. In-depth, semi-structured interviews with journalists and news media professionals add both to the exploration of automated journalism in Canada and the role and lived-experiences of journalistic actors. The interviews serve this thesis' selected hybrid thematic analysis method, which can be characterized by the application of deductive and inductive qualitative methods of thematic analysis to interpret the recorded interview data (Fereday & Muir-Cochrane, 2006). It further expands on influences drawn from Crabtree & Miller (1999) to develop a deductive codebook, while also adopting Boyatzis (1998) approach to data-driven inductive coding. In addition, this qualitative method of analysis is informed by Braun & Clark's (2006) six-step framework to conducting a rigorous and deliberate thematic analysis.

Chapter Four presents the findings from the hybrid thematic analysis, by connecting interview excerpts to the codebook developed a priori in order to highlight how the themes add to what has already been said on automated journalism. In addition, this section reviews the main findings and discusses how these intersect with the existing literature and theory, or diverge from them.

Chapter Five provides a discussion and summary of this research's main findings and reemphasizes the main objectives of this thesis and what the project has accomplished, as well as the strengths and weaknesses of the research. The findings are applied to each of this thesis' research questions in order to present a comprehensive picture of the current state of automated journalism in Canada. Components of the research methodology and the findings are further dissected with the intention of refining future approaches to research in this field. Finally, this section offers possible ways forward for the application of AI in newswork by providing an outline for the development of ethical guidelines to promote best practices for automated journalism, aimed at Canadian journalists and journalism educators.

Chapter One: Literature Review

1.1. The Human vs. Machine Cognitive Intelligence Debate

Artificial intelligence research has been around since the 1950's, with mathematicians and philosophers positing whether machines, namely digital computers, could demonstrate cognitive intelligence equal to, or surpassing, a human's capacity (Turing, 1950; Dreyfus, 1993; Haugeland, 1989). Alan Turing's (1950) essay, *Computing Machinery and Intelligence*, is an early example of the theories and future predictions of artificial intelligence's potential to compete with humans, where he predicted that before the year 2000, it would be possible to program computers to successfully play the infamous "Imitation Game." Can computers even be compared to humans? Turing, like many academics in the years following, devoted much of their work to trying to answer that question. While John Haugeland (1989) provides a Western philosophical framework of AI to demonstrate that the notion of thinking is the belief that intellection is simply the rational manipulation of mental symbols (p. 4), H. L. Dreyfus (1993) argues that AI's past failures to develop machines with mental capacities is due to this false assumption that intelligence can be reduced to mere symbolic manipulation (p. 20).

As Signorelli (2018) puts it, "the idea of [a computer] reaching and overtaking human capabilities implies the knowledge of a set of distinctive processes and characteristics which define a human being." These attributes include, but are not limited to, intelligence, language, critical and abstract thinking, the creation of art, and emotions (Signorelli, 2018). Yet, one of the most extensive areas of technological progress within the past decade has been in the advancement of artificial intelligence and its incorporation across many industries once thought to be human-driven, as was the case for the automobile industry when self-driving cars were first introduced (Lewis, Guzman & Schmidt, 2019; Diakopoulous; 2019; Broussard, 2018). Thus,

whether a computer can achieve high-level cognitive capabilities beyond the scope of a human's is a debate that is still being extensively discussed and posited to this day, including in the journalism industry. As AI-enabled technologies are becoming increasingly integrated into the newsroom, media professionals and academics are aiming to understand how it reshapes the way news is made, distributed, and absorbed by audiences, and its impacts on the current and future roles of humans in the newsroom (van Dalen, 2012; Anderson, 2013; Diakopoulous, 2019; Lewis, Guzman & Schmidt, 2019; Dörr, 2016).

However, as Graefe (2016) points out, automated news has been around for almost 50 years, initially emerging as a tool in weather forecasting. This early weather software, much like the software of today (Broussard, 2015; Lokot & Diakopoulos, 2016; Stray, 2016), functions through a series of outputs (e.g., wind speed and temperature), prioritizes by degree of importance, and generates weather forecasts using pre-determined narrative templates (Graefe, 2016). Journalists have been unconsciously supported by "less controversial" algorithms and new technologies in the newsroom for a long time, especially for editing, publishing and distributing content in broadcast production (Linden, 2017), but in recent years, the sophistication of algorithms and the rise of automated journalism have led many journalists to face its incorporation with mixed feelings, including fear and skepticism (van Dalen, 2012). However, these fears are considered to be a recent example of what Akst (2013) calls "automation anxiety," where resistance and mitigation are amplified as some journalists associate industry decline with technological advancement (Linden, 2017). According to the Canadian Media Guild, nearly 10,000 news jobs were lost in print and broadcasting between 2008 and 2013 (2013a; 2013b). In the United States, the Bureau of Labor Statistics (2018) reported that the number of working journalists decreased from 60,000 in 2005 to 44,480 in

2017. Yet, many researchers have examined to which extent technology has contributed to the current state of journalism (Deuze, 2005; Wilkinson & Winseck, 2019). Rather than attributing technology as the reason for job loss, they emphasize the need to revaluate the industry as technology continually transforms the way newswork is created and disseminated to the public (Deuze, 2005; Wilkinson & Winseck, 2019).

These varying sentiments among news media professionals have led some scholars to claim that displacement to some degree for journalists is considered inevitable with the adoption of automated journalism (Stray, 2016; Carlson, 2015). However, American computer scientist Ben Shneiderman believes these views on automation to be misguided:

> Robots and AI make compelling stories for journalists, but they are a false vision of the major economic changes. Journalists lost their jobs because of changes to advertising, professors are threatened by massive open online courses, and store salespeople are losing jobs to Internet sales people. Improved user interfaces, electronic delivery (videos, music, etc.), and more self-reliant customers reduce job needs. At the same time someone is building new websites, managing corporate social media plans, creating new products, etc. Improved user interfaces, novel services, and fresh ideas will create more jobs. (Schneiderman, in Smith and Anderson, 2014, 6)

Despite these contrasting views, some journalists have acknowledged that the development of automated journalism calls for a revaluation of the professional skills required to work in a newsroom, but still contend that creativity, personality and analytical skills remain important skills within a journalist's toolbox (van Dalen, 2012; Stray, 2016; Graefe, 2016). Still, much of the existing automated journalism research has compared machine-written content to human-produced content, where findings revealed machine-produced content to be "competitive" with that of human journalists (Graefe et al. 2018; Wölker & Powell, 2018; Lewis, Guzman & Schmidt, 2019). One study conducted in the European Union by Wölker & Powell (2018) found no significant difference in people's perceptions of human- and machine-written programs as sources, evaluating both to be credible. The debate remains over the future ability of

algorithmically-driven content to automate key journalism skills like creativity, freedom and ethical processes (Haugeland, 1989; van Dalen, 2012; Broussard, 2015).

1.2. Automated Journalism

Automated journalism can take on a wide array of meanings due to the fact that it can be applied to almost every step of the news production pipeline (Diakopoulos, 2019, p. 3; Lewis, Guzman & Schmidt, 2019; Broussard, 2015, Graefe, 2016; Dörr, 2016). In a broader sense, automated journalism can be defined as a process of news production—using structured data and algorithms—under the control of expert mechanical or electronic systems, "with little or no external influence" (Marconi, Siegman and Machine Journalist, 2017; Lewis, Guzman & Schmidt, 2019). As summarized by Marconi, Siegman and Machine Journalist (2017), these AI-enhanced programs and tools "can enable journalists to analyze data; identify patterns, trends and actionable insights from multiple sources; see things that the naked eye can't see; turn data and spoken words into text; text into audio and video; understand sentiment; analyze scenes from objects, faces, text or colours—and more."

One of the best-known applications of automated journalism is in the automatic generation of narrative news accounts using structured data (van Dalen, 2012; Carlson, 2015; Lewis, Guzman & Schmidt, 2019). Today's automated writing algorithms, like the ones developed by Narrative Science and Automated Insights (Diakopoulos, 2016; Graefe, 2016; Montal & Reich, 2017), produce news articles based on available data for topics ranging from weather and sports to finance and education (van Dalen, 2012; Graefe, 2016; Carlson, 2015). In *Automating the News* (2019), Diakopoulos provides a detailed explanation of how automated text writing works:

> The basic premise of automated text production is to take structured data, as one might find in a database or spreadsheet, and have an algorithm translate the data into written text. This process is referred to as "natural language generation" (NLG). At the simpler end of NLG are rule-based techniques that work like "Mad Libs" —that is, there are prewritten templates with gaps where numbers are dynamically inserted from a dataset according to manually crafted rules. (p. 98)

In March 2014, the *LA Times'* in-house computer program, known as Quakebot, was the first to break the news online about a 4.7 magnitude earthquake that struck Los Angeles using sensor data (Carlson, 2015; Graefe, 2016). As an earthquake occurs, the U.S. Geological Survey's Earthquake Notification Service releases an alert, and Quakebot saves a draft with all of the basic information, including the time, location and magnitude of the earthquake into the *LA Times* content management system (Graefe, 2016). Since 2014, the Associated Press (AP) has been automating corporate earnings stories and sports recaps using the Wordsmith platform developed by Automated Insights (Graefe, 2016).

Both the Associated Press and the *LA Times* had similar objectives in integrating these automated technologies, namely, producing at speed and scale. As recounted in a case study by Graefe (2016), Ken Schwencke, who developed Quakebot, awoke one morning as the earth was shaking at exactly 6:27 in the morning. Three minutes later, the story had already been published online by Quakebot, reflecting the impressive speed at which automated processes could generate a story. As for scale, automated text production allowed both media organizations to expand the quantity of news disseminated by generating articles that might not have been covered in the past due to limited time and resources (van Dalen, 2012; Carlson, 2015; Graefe, 2016). In 2015, a year after AP first introduced its automated earnings reports, it announced that the technology generated more than 3,000 stories per quarter, compared to about three hundred stories that were once written manually, thus covering the earnings of more companies than ever before (Graefe, 2016).

1.3. Algorithms in the News

To avoid confusion, it is important to reiterate that the field of Automated Journalism is not solely defined by AI-enabled methods that discount human input. Tools and content generation methods that are semi-autonomous, that is, journalistic output that is achieved through a combination of human skills and expertise with the advanced capabilities of machines, are also considered within the scope of automated journalism. The 2016 Panama Papers investigation is one of the most prominent examples of human-machine collaboration. When the International Consortium of Investigative Journalists (ICIJ) coordinated nearly 400 journalists to comb through the 11.5 million leaked documents, they harnessed the computing power of optical character optimization (OCR) algorithms in order to classify, characterize and analyze every document in detail (Diakopoulous, 2019, p. 13-14), an endeavour that may have otherwise been impossible without the help of algorithms.

Often described as algorithmic journalism, it is conceptualized here as a subfield of automated journalism, which consists of computational inputs and outputs that describe and transform data, while making inferences (Carlson, 2015; Linden, 2017; Diakopoulos, 2019). Although journalists have mixed feelings towards automated technologies (as mentioned in Section 1.2. of this chapter), many scholars emphasize the complementarity, or "hybridization" that can occur if machines assist human journalists with large-scale datasets (van Dalen, 2012; Broussard, 2015; Young & Hermida, 2015; Dorr, 2016). In a case study analysis of Narrative Science's automated journalism software and journalists' reactions to the technology, Carlson (2015) found that many of the feared predictions include "increased layoffs, polarizing personalization, and the commoditization of news writing." Whereas some scholars and journalists believe that algorithmically-driven automation tools could serve as an aid, freeing up

journalists from tedious tasks to pursue other leads and more in-depth reporting, which in turn could increase time spent on more complex investigative stories (Carlson, 2015; Broussard, 2015; Graefe, 2016; Dörr, 2016). Despite being a big proponent of algorithmic accountability—a topic which will be discussed later in this chapter—in his latest book, computational journalism expert Nicholas Diakopoulos (2019) discusses the possibilities of rapidly evolving algorithms, as they take over, or assist in, journalistic tasks:

> Algorithms are beginning to make headway in cognitive labor involving rule- and knowledge-based tasks, creating new possibilities to expand the scale and quality of investigations. Some of this technology will completely automate tasks, opening up time to reinvest in other activities. Other advances will be symbiotic with core human tasks and will, for instance, make finding entities and interpreting a web of relationships between banks, lawyers, shell companies, and certificate bearers easier and more comprehensive for the next Panama Papers. (p. 15)

Moreover, algorithms open up new possibilities in newswork, from reporting, curation, data aggregation and analysis to visualization and dissemination (Young and Hermida, 2014; Carlson, 2015; Lokot and Diakopoulos, 2015; Broussard, 2016). The benefits of automating the production of news include the low cost of integration, creating content at scale, and increasing coverage speed (Carlson, 2015; Montal & Reich, 2016; Lokot & Diakopoulos, 2016; Graefe, 2016).

While speed and scale are obvious advantages of automated news, accuracy and nuance become some of the most crucial aspects when news stories rely on the exactitude of the data rather than human oversight (Graefe, 2016; Diakopoulos, 2018). In Graefe's (2016) overview of the *LA Times* Quakebot and AP's automated earnings reports, both case studies demonstrated faults in the algorithms which led to the publication of erroneous news. In 2015, seismic sensors picked up major earthquakes that happened in Japan and Alaska, which the U.S. Geological Survey mistakenly reported as three earthquakes of high magnitude in California. Despite the

earthquakes never having actually happened in California, Quakebot gathered the data and automatically published a story for each of the three earthquakes. It was only after publication that the editors noticed the mistake and took it down from their website. That same year, AP's report about Netflix's second-quarter earnings wrongly reported that the company's share price had gone down by 71%, the reason being that the algorithm failed to understand that Netflix's stock had gone through a seven-to-one split, a financial concept beyond the scope of the algorithm's understanding. In these examples, "the human review process failed" and editors relied too heavily on the algorithms themselves (Graefe, 2016).

1.4. Accountability

Whereas the development and application of algorithms continue to grow, so has the body of literature looking into their function as knowledge producers and the subsequent impact they may have on society (Anderson, 2011; Young and Hermida, 2015; O'Neil, 2016; Noble, 2018; Diakopoulos, 2016; Broussard; 2018). Algorithms are increasingly being relied upon as a means to communicate with the public, which is why Young and Hermida (2015) coined the concept "Algorithm as Journalist," highlighting the need to deeply examine how algorithmic decisions of inclusion and exclusion are made, and whose values are being embedded into the technology. Another example of algorithmic prowess can be seen in the development of news bots. Beyond automatically disseminating news on social media platforms, news bots can also be programmed to perform a variety of tasks, such as scraping public/private websites for data, finding and identifying trending news stories and interacting directly with audiences online (Lokot and Diakopoulos, 2015; Thurman, et al., 2016; Woolley, Boyd & Broussard, 2016). One distinguishing feature of bots, as defined by Woolley, Boyd & Broussard (2016) is that as semi-

autonomous programs, “they exhibit behavior that is partially a function of the intentions that a programmer builds into them, and partially a function of algorithms and machine learning abilities that respond to a plenitude of inputs.” Returning to Young and Hermida’s (2015) notion of Algorithm as Journalist, it is essential to examine these technologies from a sociological perspective. Anderson (2013) notes that this dual, human- machine-enabled characteristic poses a challenge to journalism’s democratic function, stating that “data-crunching algorithms and other increasingly invisible information ordering devices are neither entirely material, nor are they entirely human—they are hybrid, composed of both human intentionality and material obduracy.” Moreover, it is imperative to not only study the development of news bots, algorithms, and any other (semi-) autonomous function, but also the relationships between these technologies and the people programming, or applying them to their newswork if we are to have a better gauge on these inherently programmed “intentions” and their effects on society, within the journalism industry and other sectors (O’Neil, 2016; Noble, 2018; Diakopoulos, 2016; Broussard, 2018).

Outside of journalism, algorithms are being implemented in a variety of fields such as education (Broussard; 2018; O’Neil, 2016), insurance, policing, (O’Neil, 2016; Broussard, 2016) and search engine classification (Noble, 2018). A great deal of journalism scholarship attests to the low marginal costs associated with the adoption of algorithms to generate news, and government-funded organizations have capitalized on that advantage (van Dalen, 2012; Diakopoulos, 2016; Graefe, 2016). Virginia Eubanks (2018) provides examples of its use by different levels of U.S. government, for state-funded and corporate data-mining, policy algorithms and predictive risk models, all of which are put into action for the same reasons as newsrooms: to increase productivity and to cut costs (p. 4). However, these algorithms, which

promise to boost the efficiency of classification systems like the ones in place for welfare and foster care, have been shown to "disproportionately marginalize the poor" and to make costly mistakes (Eubanks, 2018, p. 8). In her book, *Weapons of Math Destruction*, O'Neil (2016) finds that algorithms, though often created with good intentions, are "opaque, unquestioned and unaccountable" (p. 12), urging researchers to explore ways to hold algorithms accountable.

Algorithmic accountability, a method of verification that has been referred to by many experts as an "algorithm audit" (O'Neil, 2016, p. 208; Broussard, 2018, p. 194; Diakopoulos, 2019, p. 207), presents itself as a way of providing descriptions, explanations and justifications (Dörr & Hollnbuchner, 2017) for "the behaviour of decision-making algorithms, particularly in cases where there was a fault or error" (Diakopoulos, 2019, p. 207). Harvard Law professor Lawrence Lessig writes (in Broussard, 2018), "if code is law, then we need to make sure the people who write code are doing so in accordance with the rule of law" (p. 194). However, no agreed-upon algorithm-design laws and verification processes have been established to date, which proves to be a challenge if we are to hold algorithms accountable for their actions (Broussard, 2018). Still, efforts toward designing and auditing algorithms are already being researched for their societal and economic impacts at Princeton University, the Massachusetts Institute of Technology, New York University, and other academic institutes worldwide (O'Neil, 2016, p. 210; Broussard, 2018, p. 195). In the journalism sector, a sub-field of computational journalism, known as algorithmic accountability reporting (Broussard, 2018, p. 195; Diakopoulos, 2019, p. 205) has emerged in an attempt to investigate and better understand how algorithms are designed and in what ways the media can hold them accountable while fulfilling their watchdog role in the public interest.

Despite these efforts, little research has looked at how to uphold journalistic ethical standards when newsroom algorithms are implemented. Diakopoulos (2019) stresses that algorithms that perform newswork must also be held to account, as they have become increasingly integrated for their "data mining, automatic content production, newsbots, and distribution chain" capabilities (p. 234). In journalistic practice, algorithms now exercise editorial decision-making powers that can choose what information is included or excluded from a story, and how that story will be disseminated and perceived online (Annany, 2016). Inherent biases built into algorithms can "skew public perceptions or guide people astray in ways that undermine democratic ideals," which is why journalism transparency and ethical guidelines need to be put in place to accommodate these technologies (Diakopoulos, 2019, p. 234).

The critical need for established transparency and accountability guidelines in automated journalism has become an urgent topic of discussion (Broussard, 2016; Diakopoulos, 2016; O'Neil, 2016; Eubanks, 2018; Wolley, Boyd & Broussard, 2016). Diakopoulos (2016) asks the question: "In the face of important or expensive errors, discrimination, unfair denials of public services, or censorship, when and how should algorithms be reined in?" The full disclosure of information and authorship for any news story has always been a vital part of journalism practice in serving the public interest (Montal & Reich, 2017). As such, salient disclosure policies and ethical guidelines must be reviewed, as algorithms and AI-enabled technologies increasingly encroach the journalism landscape in Canada and elsewhere. As Diakopoulos (2019) points out:

> The main challenges are how to harness the power of algorithms to speed up, scale up, and personalize content while staying true to preferred values, and how to steer toward outcomes that are beneficial to individual groups, and society while balancing business concerns. [Newsrooms] need to be more deliberate about the values they are building into systems, particularly if they want those values to be reflected prominently in the algorithms and artificial intelligence that drive the future of media. (p. 241)

The challenges of algorithmic authorship have been explored by Montal & Reich (2017), whose findings showed discrepancies in the practices concerning bylines and full disclosure policies of media organizations implementing automated journalism. Montal & Reich's (2017) research on algorithmic authorship reveals the critical need for more comprehensive and consistent algorithmic transparency alongside media organizations' bylines. Thus, disclosing the "methodology,[1] construction and limitations of the algorithm" (Montal & Reich, 2017) should be an integral and mandatory part of a newsroom's ethical practice (Broussard, 2016). However, no such research has been conducted extensively in Canada, which makes this present research important in order to fill the knowledge gap.

Furthermore, Woolley, Boyd & Broussard (2016) suggest that bots should be thought of as a "civic prosthetic" that supercharges reporting practices by automating tasks, particularly in terms of data analysis to expose interesting connections and patterns in data, which would typically take human reporters' hours to achieve. As a civic prosthetic, automated journalism then retains the added necessity of human input to verify automated outputs, thus creating a bot-human reporter relationship that augments journalistic content (Broussard, 2015; Graefe, 2016; Stray, 2016; van Dalen, 2012). An example of this integration is mentioned by Stray (2016) for the AI app Wibbitz, a software that creates video rough cuts from texts, which are verified and edited by a human afterwards. The value of an AI for journalism like Wibbitz, Stray (2016) mentions, is "more about the workflow and less about sophisticated AI."

Yet, automated journalism software does put pressure on the performance of human reporters, as they increasingly compete with machines that can produce news at a scale which is

[1] In this context, disclosing a methodology refers to divulging what statistical methods were used for mining and analyzing data for any given story, which would allow the reader to assess the steps undertaken by an algorithm to reach a defined outcome (Diakopoulos, 2014; Linden, 2017)

unmatched by humans (van Dalen, 2012; Stray, 2016; Lokot & Diakopoulos, 2016). Moreover, it is crucial to conduct qualitative research in order to gain insight into the shifts in skills and requirements to remain employable and competitive in this emerging field.

Chapter Two: Theoretical Framework

Journalism and technology have always been deeply intertwined in complex and diverse ways. From the telegraph to the typewriter, technological developments have had significant impacts on the proliferation and sophistication of news production, distribution and consumption (Anderson, 2012; Zamith & Braun, 2019; Lewis & Westlund, 2016). As Diakopoulos (2019) puts it, "the history of journalism is one of adaptation as news technologies—telephony, photography, reproduction, and computerization—changed the nature of roles, tasks and workflows" (p. 5). Journalism has been viewed by many scholars to be in continual ideological development as factors like commercialization, bureaucratization and emerging media technologies repeatedly call for a revaluation of the values, ethics and characteristics of quality journalism needed in order to sustain the profession (Deuze, 2005; Anderson, 2013; Lewis & Westlund, 2015).

Yet, technology has more often than not been studied for its societal impacts as the medium or channel of human-to-human interaction, rather than as the communicator itself (Guzman, 2018). If contemporary journalism is to be better understood as new forms of technology make their way into the news pipeline, while no longer only serving as the communication medium, research cannot exclusively be focused on journalistic actors and their work and values, as this would provide knowledge on merely one aspect of the news media organization. With that in mind, this thesis draws from Lewis and Westlund's (2015) call for a more holistic approach, by attempting to understand the diverse relationships between journalism and technology, which encompasses journalists, technologists and technological actants (e.g., algorithms, networks, computer management systems). On the other hand, it is important to note the distinction between journalists and technologists. Journalists, as described by Lewis and

Westlund (2015), involve human actors encompassing the roles of reporters, editors, producers, and others who are crafting output identified as news, while the authors describe technologists as "information technology specialists, systems designers, project managers, information architects, product developers, and other programming technicians," and focus on the relationships between journalists and technologists in newsrooms (Lewis and Westlund, 2015). However, for the purpose of this thesis, technologists are also considered within Nikki Usher's (2016) definition of interactive and data journalists, as they are individuals who know how to build code, apply algorithms and sophisticated software to produce a journalistic product relevant to the public interest (p. 24).

Technological actants arise out of Lewis and Westlund's (2015) "Four A's" model, and are defined as nonhuman, material objects that interact or intersect with humans and human endeavors (Zamith & Braun, 2019). Although Lewis and Westlund (2015) draw upon Latour's Actor Network Theory, mainly in their acknowledgement of the idea that actants and actors have significant relationships that must be studied to better understand our world, they shy away from Latour's model by arguing for the need to distinguish nonhuman and technological actants from human and social actors. Consequently, they seek to understand whether technological actants make a significant difference to the activities of journalists and technologists in the newsroom. As such, Lewis and Westlund's (2015) model aligns with each of this thesis' research questions. Firstly, their framework can help us better understand who produces automated journalism in Canada on an individual (journalist) and organizational (newsroom) level, because it explores the hybrid arrangements among journalism actors in accordance to those automated technological affordances that support news production and distribution. Secondly, the "human-technology dimension as an organizing framework" allows us to investigate the type of news content

produced and to what capacity, while also gaining insight into the realities of the journalists and editors working alongside automation. By applying this framework, we can recognize the interplay of automated processes and journalists/editorial workers and how they are "negotiating issues of authority, identity and expertise" as these technologies assume more responsibility for newswork traditionally accomplished by a human actor (Bakker, 2012; Lewis, 2012; Westlund, 2011, in Lewis and Westlund, 2015).

A growing field of study, known as Human-Machine-Communication (HMC) has emerged in response to new technologies that serve the function of "message sources" rather than "message channels" (Lewis, Guzman & Schmidt, 2019; Guzman, 2018). HMC is defined by Guzman (2018) as "the creation of meaning among humans and machines and the study of this meaning-making and related aspects" and aims to develop a theoretical framework for human-to-machine communication, while also shedding light on the social and cultural implications of such technologies as they take on communicative roles once solely reserved for humans (p. 3; Lewis, Guzman & Schmidt, 2019). Moreover, informing this thesis—and in support of Lewis and Westlund's (2015) model—are ideas borrowed from HMC that provide a better understanding of the relationships that are formed between news professionals and automated technologies, and what social and ethical implications they portend for the future of newswork (Guzman, 2018; Lewis, Guzman & Schmidt, 2019). HMC also helps elucidate the underlying motives that led journalism practitioners to adopt and engage with these technological tools in the first place (Guzman, 2018, p. 27).

This research identifies automated journalism as an area of computational journalism, which is considered to be omnipresent forms of algorithmic, social scientific and statistical newswork (Anderson, 2012; Diakopoulos, 2015) that are increasingly integrated into journalistic

practices. While computational journalism has been defined as a way of using big data for help in newsgathering and storytelling (Anderson, 2012; Carlson, 2015), automated journalism is referred to here as the process of using sophisticated software to interact online and to generate and find news content with little-to-no human input—apart from the initial step of programming the algorithm (Graefe, 2016; Broussard, 2015; Thurman et al., 2016).

Within the scope of this research, it is important to note that automated journalism is considered within Lewis and Westlund's (2016) fourth typology of technological dependence within journalism, namely *Technology-oriented journalism*. It is defined as a type of symbiotic relationship between human actors and nonhuman actants that vary in terms of their interdependence on one another. Since automated journalism has the ability to carry out newswork under limited or no supervision of technologists and/or journalists, it becomes crucial to consider labour shifts in news production (Lewis and Westlund, 2016).

This thesis project considers the potential roles and limitations of AI and automation in Canadian journalism within an industry that is constantly competing to turn up captivating and original story ideas in today's digital age (Broussard, 2015; Lokot & Diakopoulos, 2016). One of this project's considerations involves the tensions that emerge between humans and technology as the two become increasingly interdependent (Lewis & Westlund, 2016), as this will help provide an overview of the journalistic roles and skills required to be a working journalist within a media landscape that is continuously changing to make space for new sophisticated technological tools. Automated journalism has been previously discussed as a way to enhance a newsroom's productivity and efficiency (Broussard, 2015, Stray, 2016) with structured-data reporting (Graefe, 2016; Dörr, 2016; Stray, 2016).

Furthermore, as news organizations continue to carve out space for automated journalism, its ability to take over tedious tasks once performed solely by journalists (Graefe, 2016; Carlson, 2015; Stray, 2016) highlights the need to focus on technology not just as the communication medium, but as the communicator. While the development and integration of AI-enabled communication technologies has led to a more sustained effort to understand the ways in which the flow of information is transformed (Pavlik, 2000), more research needs to be done in order to understand the diverse and complex relationships that emerge when technological actants assist journalists in content production, take over some journalistic roles, and give rise to new areas of journalism, such as algorithmic accountability reporting. As such, Lewis and Westlund's (2016) holistic approach and Guzman's (2018) Human-Machine-Communication framework will help address who is exercising automated journalism in Canada, where these technologies situate themselves in the news production pipeline, how automation impacts current and future structures of newswork, and whether the skills and values required from journalists will need to be reoriented. By applying this theoretical lens to the lived experiences of Canadian journalists working alongside algorithms (see Chapters 4 & 5), this thesis hopes to contribute knowledge about the impact automated journalism's increased presence has on Canadian journalistic practice and to provide useful information about journalists working alongside these technologies to inform and bolster newsroom practices and journalism education.

Chapter Three: Methodology

This chapter outlines the methodology selected in order to answer this thesis' main research questions (see Introduction, p. 3-4), through interviews and a subsequent hybrid thematic analysis of the interview transcripts. Since little is known about the extent to which automated journalism has been adopted in Canada, and the realities of the journalists working alongside it, in-depth semi-structured interviews add both to the exploration of the current state of automated journalism in Canada and the lived experiences of journalistic actors. In-depth semi-structured interviews allow the "descriptions of the lived world of the interviewees with respect to interpretations of the meaning of the described phenomena" (Kvale, 1996, p. 30). The purpose of this type of qualitative interview is to contribute to a body of knowledge that is both theoretical (informed by existing literature), which guides the interview protocol, and conceptual in that it explores "the range of attitudes, perceptions, beliefs and behaviours" of the participants in order to gain insight on the object of study (Di-Cicco & Crabtree, 2006; Roberts, Dowell & Nie, 2019). Moreover, the open-ended nature of the questions helps define this thesis' research problematic while also leaving room for spontaneity in the answers offered by the participants, therefore eliciting rich data.

In order to analyze the data collected from the in-depth semi-structured interviews, this thesis resorts to a hybrid thematic analysis approach informed by the work of Fereday & Muir-Cochrane (2006). It involves the creation of a codebook based on a theory-driven deductive approach to help guide the data analysis (Crabtree & Miller, 1999), followed by Boyatzis' (1998) data-driven inductive approach of thematic analysis, allowing themes to emerge directly from the interview data. This combined deductive/inductive approach was chosen as it allows for a more

complete analysis of the data in order to be best able to answer the research questions while demonstrating rigour (Fereday & Muir-Cochrane, 2006; Roberts, Dowell & Nie, 2019).

3.1. Participant Sampling

Since the existing academic literature on the topic suggests that automated journalism is still in its "early-market phase" (Graefe, 2016), and no known research has explored to which extent media organizations and professionals are using automated journalism in Canada, identifying a list of those that do required some online searching. The initial recruitment stage was informed by extensive online research through Google's search engine to explore topics covering any newsroom adoption of automated journalism in Canada. The Canadian Association of Journalists' website and social media sites, primarily Twitter and LinkedIn, were also resources used for recruitment. Additional recruitment was made possible through connections made through Concordia University's Department of Journalism. Further sampling based on suggestions offered by study participants, or recruited individuals who chose to not participate, provided the possibility to expand the search for more recruits that fit the participant criteria. This type of snowball sampling served as a means to advance the initial recruitment stage, particularly by serving as an aid to identify unknown participants who could significantly contribute to the ongoing research and thus, add knowledge value (Browne, 2005).

Several criteria delimited the selection process of candidates who would be asked to participate in the in-depth semi-structured interviews. An interview sample logically depends on a researcher's goal of pursuing a study and its intended outcomes. Given this thesis' research questions, a purposeful sampling strategy was used, as this sampling technique can inform an understanding of a research problem or describe a central phenomenon through the lived experiences and detailed views of people in this field (Creswell & Poth, 2018, p. 125). Following

the initial online search phase, recruitment began by targeting journalists and programmers working for English- and French-language news media outlets known for their automated journalism production through their reporting, analytical use or social media presence. As mentioned in Chapter 2, programmers who work in newsrooms were considered for recruitment based on Usher's (2016) assertion that programmers who are fluent in code are in essence journalists, because they must know "how to communicate in a newsroom workflow, build products for acts of journalism, and are integrated into editorial workflow" (p. 98). In addition, professors from journalism departments in universities across Canada were considered if they not only taught automated journalism methods in their curricula, but also currently worked as journalists.

The final list of potential study participants included 26 individuals from 10 media organizations, four Canadian universities, and one freelance journalist. All of the media professionals on that list were contacted via email between October 2019 and February 2020.

3.2. Interviews with Key Figures

In total, 10 key figures from four media organizations and two Canadian universities in Quebec and Ontario agreed to be interviewed between October 2019 and March 2020. One journalist had to be excluded from the data analysis because she had no experience in the Canadian sector, but segments of her interview will be included in the discussion section (Chapter 5) to contextualize larger issues in the automated journalism field. As such, nine news media professionals were interviewed for this study, and each of their professional identities can be broken down as follows: one digital news editor and content strategist, the head of a digital data desk, the vice president of a digital data desk, one interactive graphic journalist, three data

journalists and two working data journalists who also teach university-level journalism courses (See Table 1 for a detailed summary of each participant). The initial goal of this thesis was to conduct 12 to 15 interviews to ensure that enough data was collected for this specific project, which made the total of nine interviews less than ideal. However, given the limited knowledge on the current state of automated journalism in Canada, and that the field is relatively young, this is not an unreasonable response rate. However, future research may require more interviews to fully saturate the model.

Table 1. Description of Research Participants

Participant ID	Gender	Profession	Location (Province)[b]	Initial Recruitment Date	Date of Interview	Interview Method
Journalist 1	M	Digital News Editor and Content Strategist	Quebec	October 14, 2019	October 24, 2020	In person
Journalist 2	M	Data Journalist	Quebec	October 8, 2019	October 24, 2020	In person
Journalist 3	M	Data Journalist	Quebec	October 8, 2019	November 11, 2020	In person
Journalist 4	M	Head of Digital Data Desk	Ontario	October 15, 2019	November 15, 2020	Skype
Journalist 5	M	Vice President, Digital Data Desk	Ontario	January 28, 2020	January 31, 2020	Phone
Journalist 6	M	Interactive Graphic Journalist	New York[c]	January 27, 2020	February 14, 2020	In person
Journalist 7	M	Data Journalist and Director	Ontario	January 27, 2020	February 20, 2020	Phone
Journalist 8	M	Professor and Data Journalist	Quebec	February 25, 2020	February 26, 2020	In person

Journalist 9	M	Data Journalist	Ontario	February 25, 2020	March 4, 2020	Phone
Journalist 10[a]	F	Data News Developer	New York	January 27, 2020	February 14, 2020	In person

a. Journalist 10 was excluded from the data analysis because she works for an American news media organization and has no experience in the Canadian sector.
b. For confidentiality reasons with respect to this thesis' ethical guidelines, only the provinces where the participants are currently employed are disclosed.
c. Despite working for an American news outlet, Journalist 6 was still considered for the purpose of this thesis as this participant is a Canadian citizen who worked with automation tools at two different Canadian media organizations before moving to the U.S.

Interviews were conducted in person, over the phone or through Skype (video-camera enabled communication platform) depending on the participant's availability, location, and other logistical factors. All of the participants provided written and verbal consent to participate before the interviews were conducted (Concordia University ethics certification number: 30011687). They were asked a focused set of questions which aimed to understand how their experiences, skills, practices and opinions of automated journalism have impacted their jobs and newsrooms, as well as their perceived notions on the necessity or optionality of automated journalism in today's news environments. Interviews lasted between 45-90 minutes and were audio-recorded. Additionally, the interviews were conducted in either French or English, depending on the participant's preferred spoken language. Their replies were transcribed and systematically categorized into recurring themes using a hybrid thematic analysis method (Fereday & Muir-Cochrane, 2006), which will be discussed in detail in the next section of this chapter.

The interviews with news media professionals were conducted following an initial interview protocol of 13 questions. However, the protocol was subject to change as a semi-structured interview guide, "no matter how extensive its preparation, should still be considered a work in progress" (Adams, 2015, p. 499) as interviews are conducted and unanticipated themes emerge. The first four participants brought up several themes (data privacy/security, defining

automated journalism, language as an adoption barrier) which were not addressed in the original interview script and that were deemed relevant for the purpose of this research. This prompted the addition of three more questions to ensure that all important themes were covered in the subsequent interviews. The first four interview participants were contacted by email in order to get their written responses to the added questions which were not directly addressed during their interviews.

As such, the final interview script was based on the following 16 guiding interview questions: **(1)** What does being a journalist mean to you? **(2)** In your opinion, are journalists generally receptive of new technologies in the newsroom? Why or why not? **(3)** How would you define automated and semi-automated journalism in the context of your work? **(4)** What types of algorithmically driven or automated tools and processes do you use in your newsroom? **(5)** When did you begin adopting these types of journalism tools and what were your motives behind adopting them? **(6)** What types of stories or tasks does automated journalism help you tell or accomplish? **(7)** How has the application of automated journalism in your newsroom changed your job? **(8)** How has automated journalism, if at all, shifted journalistic labour and skills? **(9)** Do you view automated journalism as a benefit or threat to your profession? More specifically, a threat to those who don't practice it? **(10)** How, if at all, has the use of automation and algorithmically-driven tools changed authorship and transparency guidelines? **(11)** Do you think newsrooms should adopt policies/strategies surrounding these technologies? Why/why not? **(12)** In your opinion, do we need to establish data journalism ethics around the process of scraping or downloading data from non-government websites? Why/why not? **(13)** How are the algorithms driving the content and stories produced held accountable? **(14)** In your opinion, what are the advantages and disadvantages of using these tools to communicate and interact with your

audiences? **(15)** Based on your knowledge, do you think Canada is behind in adopting automated journalism? Why/why not? **(16)** Do you think the existing language barrier between Quebec and the rest of Canada makes it difficult to train people in this field of journalism?

The interview questions listed above were largely informed by an extensive review of the literature on automated journalism, algorithms in the newsroom and AI-enabled journalistic methods. As aforementioned in Chapter 1, the literature demonstrated many different ways to exercise automated journalism in order to produce journalistic output (Diakopoulos, 2019, p. 3; Lewis, Guzman & Schmidt, 2019; Broussard, 2015, Graefe, 2016; Dörr, 2016). The literature also addresses journalists' sentiments towards automated journalism (van Dalen, 2012; Carlson, 2015; Stray, 2016; Graefe, 2016), and questions regarding ethics, transparency and accountability as these technologies increasingly take on journalistic roles (Anderson, 2011; Young and Hermida, 2015; O'Neil, 2016; Montal & Reich, 2017; Noble, 2018; Diakopoulos, 2016; Broussard; 2018). Since no previous study has looked at what this means for Canadian journalism specifically, these open-ended questions were designed to get a better understanding of the current state of automated journalism in Canada.

All of the participants showed a deep understanding of, and engagement with, the implementation of automated journalism and provided a rich range of insights and perceptions of motivations behind the adoption of AI-enabled communication technologies, as well as the domains of coverage (politics, sports, business, etc.) and the various tools involved (i.e., content generation software, scraper bots, automated tracking). Due to their variety of roles, the participants offered reliable testimony on transparency and authorship, ethical considerations and implications for journalists working alongside automated journalism, as well as the future of newswork in Canada.

3.3. Hybrid Thematic Analysis

To analyze the data from the in-depth semi-structured interviews, this thesis utilizes a hybrid thematic analysis method informed by the work of Fereday & Muir-Cochrane (2006), which relies on a dualistic technique of inductive and deductive thematic analysis. Thematic analysis is a method for identifying, analyzing, organizing, defining, and reporting the emergence of themes within a data set that are valued by the researcher as being important to the description of the object of study (Fereday & Muir-Cochrane, 2006; Braun & Clarke, 2006; Nowell, et al., 2017). Fereday & Muir-Cochrane (2006) write that it is a "form of pattern recognition within the data, where emerging themes become the categories for analysis." Boyatzis (1998) defines it not as a specific method, but as a tool to use across different qualitative methods, which is a sentiment echoed by Ryan and Bernard (2000), who locate the action of coding in thematic analysis as a process founded in analytic traditions, such as grounded theory, which is concerned with "the discovery of theory from data systematically obtained and analyzed in social research" (Glaser & Strauss, 2002, p. 12). On the other hand, Braun & Clarke (2006) believe that thematic analysis should be considered a specific method, arguing that "through its theoretical freedom, thematic analysis provides a flexible and useful research tool, which can potentially provide a rich and detailed, yet complex account of the data" (p. 78).

Yet, it is important to note that within qualitative research, there is a lack of texts and resources on rigorous and relevant approaches to thematic analysis, which has significant implications in terms of the credibility of the method, which is why many researchers in recent years have attempted to demonstrate its value as a method not tied to a particular epistemological or theoretical perspective (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006; Maguire &

Delahunt, 2017, Nowell, et al., 2017). As such, this thesis works toward rigour and replicability using Fereday & Muir-Cochrane's (2006) hybrid approach of inductive and deductive coding and theme development, while also drawing from Braun & Clarke's (2006) six-step framework for conducting thematic analysis in a "deliberate and rigorous way." It should be noted that the fields both these thematic analysis methods originate from may not directly connect to journalism. Fereday & Muir-Cochrane's hybrid analysis stems from health and nursing, while Braun & Clarke apply their method to psychology. Thematic analyses from these specific fields, while very different from the present study, are useful to journalism research, as this method is not so much about the study of distinct professions, as it is about understanding the lived experiences of the practitioners who work in these fields.

Although most thematic analyses in qualitative research opt for one coding method (either inductive or deductive) over another, this thesis' twofold approach to thematic analysis incorporates both Boyatzis (1998) data-driven inductive approach and the deductive approach outlined by Crabtree & Miller (1999) for the creation of a preliminary codebook (Fereday & Muir-Cochrane, 2006).

The deductive component, known as a top-down approach (Braun & Clarke, 2006) is concerned with the development of a codebook as a point of departure, using predefined themes grounded in this thesis' theoretical framework, which will be discussed in the next section of this chapter. The interpretive—or bottom-up—approach followed the creation of the codebook, allowing themes to emerge from the answers to the interview questions, which were then categorized as subcodes (Maguire & Delahunt, 2017). Fereday & Muir-Cochrane (2006) emphasize this hybrid approach as complementary to the research questions "by allowing the tenets of social phenomenology to be integral to the process of deductive thematic analysis while

allowing for themes to emerge directly from the data using inductive coding." In *Learn to build a codebook for a generic qualitative study* (2019), Mihas and the Odum Institute write the following:

> A study that employs both deductive and inductive codes invites researchers to carry with them what they know from the literature and also listen to what is "new" in the data, possibly problematizing earlier assumptions or logics. In this case, we might find that concepts in the literature are interpreted differently by the participants themselves. A cancer survivor who calls himself a "cancer graduate" might invite us to rethink the medical definition of "survivor." (p. 4-5)

Although the coding process for each interview takes a systematic, iterative approach, it involves recognizing an important moment in the conversation and encoding it prior to proceeding to the process of interpretation (Boyatzis, 1998; Fereday & Muir-Cochrane, 2006). A "good code," as suggested by Boyatzis, is one that captures the qualitative richness of the object of study. The encoding process of the data helps to identify and develop themes. Boyatzis defines a theme as "a pattern of information that at minimum describes and organises the possible observations and at maximum interprets aspects of the phenomenon" (p. 161).

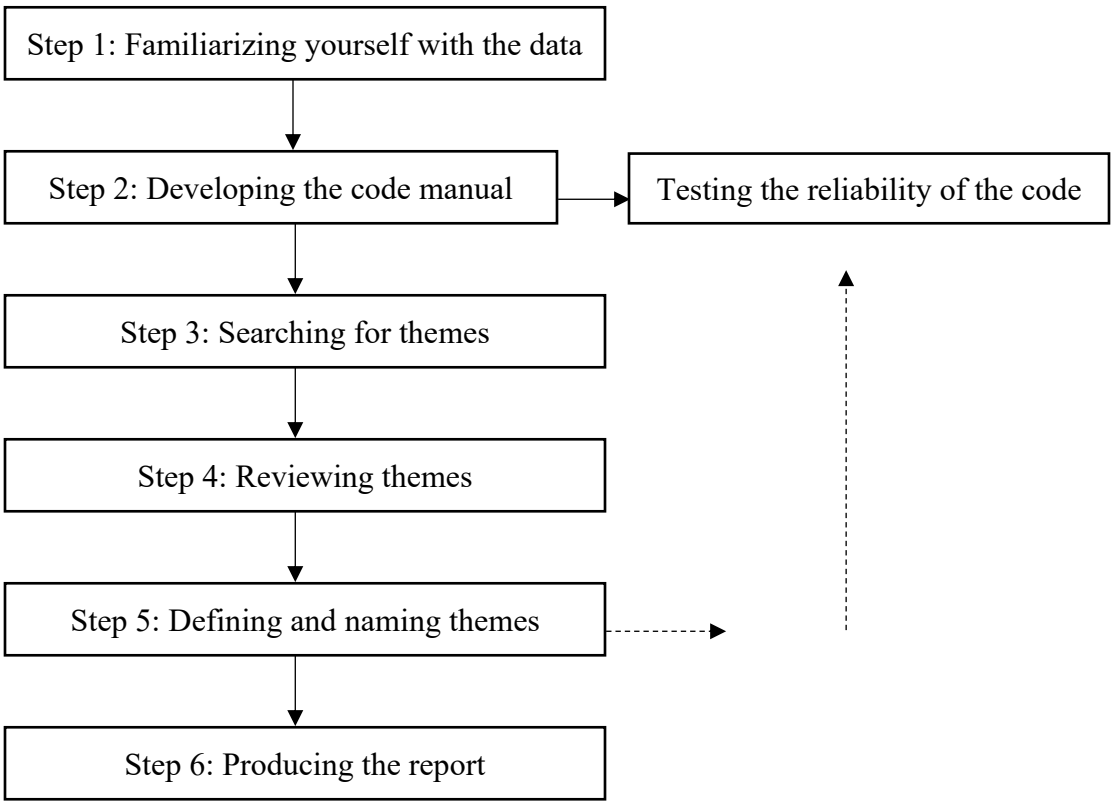
A comprehensive process of data coding and identification of themes from the interview transcripts was undertaken with the use of Excel in order to facilitate data management. This systematic, step-by-step process is outlined below. Although presented as a linear procedure, it is important to emphasize that analysis is more of a recursive process, where the iterative analysis allows for movement between the phases as needed (Braun & Clarke, 2006). This fluid process—or interactivity—of qualitative inquiry, is described by Tobin and Begley (2004, in Fereday & Muir-Cochrane, 2006), as the overarching principle of "goodness." According to Fereday & Muir-Cochrane (2006), "goodness" is achieved by conducting the data collection and analysis stages concurrently, and by repeatedly reading the thematic analysis stages guiding one's study before moving forward with additional analysis, to ensure that the developing themes

are “grounded in the data.” The primary objective of data collection in this current study was to help determine the present state of automated journalism in Canada and to provide a framework for future research into the development of potential ethical guidelines for Canadian journalists and journalism educators as automation continues to find its way into the newsroom.

3.4. Developing the Coding Process

The chart shown in Figure 1 represents each stage of coding based on Braun & Clarke’s (2006) six-step approach to conducting a rigorous thematic analysis, while also including the dualistic inductive/deductive approach from Fereday & Muir-Cochrane (2006).

Figure 1. Diagrammatic representation of the stages undertaken to code the data (adapted from Braun & Clarke, 2006, and Fereday & Muir-Cochrane, 2006)



Once the interview data is collected through transcription, the most straightforward step to thematic analysis involves familiarizing oneself with the data through repeated reading of it in an "active way," meaning the researcher is continually immersing himself/herself in the data with the goal of searching for, and identifying, meanings and patterns (Braun & Clarke, 2006). Although this step may come with prior knowledge of the data (i.e., based on existing literature) and a researchers' initial interests or thoughts, it is a vital step of immersion in iterative qualitive methodologies (Braun & Clarke, 2006). While Braun & Clarke describe the second step as "generating initial codes," they point out that this phase will depend on whether one's analysis is theory-driven or data-driven. In the former, Braun and Clarke (2006) suggest that a researcher may approach the data with pre-defined questions that inform the coding process, but in the latter, the themes depend on the data. Since this phase of their six-step thematic analysis guide does not consider the possibility of employing a dualistic inductive/deductive technique, this step has been modified to incorporate the development of the coding manual informed by the work of Fereday & Muir-Cochrane (2006). As such, "generating initial codes" is replaced with this thesis' initial step of developing the coding manual and assessing its reliability. According to Fereday & Muir-Cochrane (2006), the generation of a codebook strengthens the credibility of a study by providing readers with a "clear trail of evidence" that demonstrates how data was interpreted, even if inductive coding is to follow. The development of the codebook was an important initial step, because it served as a data management tool in its own right, thus allowing data to be organized in segments of similar or related text to assist in the subsequent interpretation of the raw data (Crabtree & Miller, 1999).

As such, this thesis' theoretical framework, namely, exploring the diverse relationships between news professionals and automated technologies (Lewis & Westlund, 2015; Guzman,

2018), served as the starting point for the a priori development of the coding template. The codebook is therefore composed of four main codes (self, self in relation to technology, collective self in relation to technology, and technology as the professional self), which were subsequently divided into 10 subcodes (see **Table 2**).

Table 2. Codebook for analysis

Code	**Code Description**	**Subcodes**	**Description of Subcodes**
1. Self	This code describes the participant's professional/personal identity, which informs how they interpret their role as a journalist (communicator)	a. Personal Identity b. Professional Identity	a. Describes how participants define themselves on a personal level b. Outlines the participants view of the purpose they serve in their professional role with regards to being a journalist
2. Self in relation to technology	The relationship between participants and automated technologies with respect to how they interact, how it has changed their job and their general considerations and observations on the extent to which the relationship is reciprocal	c. Relationship with technology d. Level to which technology has changed their job e. Level of reciprocity	c. Describes how participants' view their relationship with technology with respect to their professional experience d. The extent to which AI-enabled technologies have shifted their professional role e. The extent to which technologies assist (in part) in the participants' newswork, and the types of tasks they accomplish and the stories they produce

3. Collective self in relation to technology	Outlines the respondents' observations of the relationship between the collective self (representation of self at a group level) and automated journalism from a historical, economic and social perspective. It also outlines motives for which the collective self would adopt automated journalism	f. Reception of technologies within media organizations over time g. New technologies and the journalistic business model h. Motives and limitations	f. Explores the relationship of technology in relation to the collective self from a historical and socio-cultural perspective g. Describes the economic context of the emergence of automated technologies and the participants' opinions and perceptions of its necessity h. Evaluates the reasons given by the participants as to why (or why not) Canadian media organizations would choose to adopt automated journalism, while also discussing the limitations
4. Technology as the professional self	Situates the socio-cultural implications of the emergence of technology as the journalistic actor (communicator) and outlines the way technology shifts current journalistic standards in terms of policy, while also exploring audience perception	i. Ethical implications j. Audience engagement	i. Describes the participants' opinions and perceptions of technology as communicator and the ethical implications j. Evaluates the participants' perceptions of audience engagement with regards to technology as the communicator

Following the development of the coding manual (**Table 2**), an essential step in assessing its reliability as a useful framework is to determine its applicability to the raw data (Boyatzis, 1998; Fereday & Muir-Cochrane, 2006). Data from the first interview conducted with Journalist 1 (see **Table 1**) was selected as the trial piece and was applied to the theory-driven codes. The resulting coding process was shared with Dr. Elyse Amend, this thesis' supervisor, for review. An example of applying the template of codes to the data collected from Journalist 1 is given in Table 3.

Table 3. An example of coding with the template of codes developed a priori.

Theory-driven code	**Data from interview with Journalist 1**
Self in relation to technology	"The way it's changed my job is that I went to get a certification in product management just because I hadn't done any product management before. Now that I oversee the creation of AI products for the newsroom, I wanted to make sure that things were getting done properly" "I'm the digital news editor, so what I still do is assist reporters in telling their stories in a way that is clear for an online audience. Now I need to ask myself, 'Ok, how do we make sure that we produce a tool that has a benefit for everyone in the newsroom?' and that wasn't a natural step for me."
Reflection	Acquiring new skills: AI product management and development Self-assessment—acknowledging that automation tools change what is required of self
Intention and Outcome	Goal to develop technologies that benefit professional self and collective self

Testing the reliability of a codebook is a crucial step to qualitative inquiry, since the guiding framework must ultimately preserve the participant's subjective point of view and acknowledge the context within which a given phenomenon is studied (Fereday & Muir-

Cochrane, 2006). This form of interpretive rigor must demonstrate clearly how the researcher achieved a deep understanding of the data, and should present supporting quotations from the data (Rice & Ezzy, 1999; Fereday & Muir-Cochrane, 2006). In Table 3, Journalist 1's comments are conveyed in his own words with direct quotations in order to demonstrate how the overarching theme, 'Self in relation to technology,' is supported by the raw data. The subsequent rows, labelled 'Reflection' and 'Intention and Outcome' present the researcher's interpretation, or conclusions, that were drawn from the statements conveyed by Journalist 1 in relation to the theme, thus evaluating their credibility and relevance to the phenomenon under analysis. Following Dr. Amend's review of the sample data from Journalist 1 and its applicability to the predefined codes, no modifications to the codebook were required.

Step three, according to Braun & Clarke (2006), begins when all data have been initially coded and collated, thereby refocusing the analysis and sorting different codes into potential themes and collating all relevant coded data extracts within identified themes. This step is important as it provides a preliminary collection of "candidate themes, sub-themes, and all extracts of data that have been coded at this point" (Braun & Clarke, 2006). For this research project, this step also involved using the template analytic technique (Crabtree & Miller, 1999), which involved applying the codebook to the text while performing the iterative inductive analysis. Analysis of the data at this stage was guided, but not confined, by the codebook. During this phase of analysis, inductive codes were assigned to segments of data that presented a new theme from the interviews (Boyatzis, 1998; Fereday & Muir-Cochraine, 2006). These additional codes, as exemplified in Fereday & Muir-Cochrane's study, and echoed in this thesis' hybrid analysis, were either separate from the predetermined codes or expanded a code from the coding manual. Step four, reviewing the themes, involved reviewing all the collated extracts for each

theme and deciding whether they appeared to form a coherent pattern (Braun & Clarke, 2006). This was then followed by reviewing the themes in relation to the entire data set and determining whether the candidate themes accurately reflect the meanings of the data set as a whole (Braun & Clarke, 2006). 'Accurate representation' was achieved when the themes, collected both inductively and deductively, were clearly understood in how they fit together and how they tell the overall story about the data (Braun & Clarke, 2006). As Braun & Clarke (2006) point out, coding and generating themes could go on *ad infinitum* as the opportunities for coding qualitative research are endless. Moreover, the coding of this thesis' data was considered complete when existing codes started to repeat themselves within the interviews. Once no new codes emerged, the process was considered complete.

Finally, the fifth step before producing the findings involved defining and further refining the themes that will be presented in the findings section. More specifically, this step concerns identifying the meaning behind each theme and determining what aspect of the data each theme captures, and whether certain themes contain sub-themes (Braun & Clarke, 2006). At this point in the analysis, the researcher should be able to clearly identify the "essence of what each theme is about" and form a coherent and consistent narrative that not only identifies the 'story' that each theme conveys on their own, but the broader overall story of the data in relation to the research questions (Braun & Clarke, 2006).

Chapter Four: Findings

As mentioned in the previous chapter, the codebook (**Table 2**) was informed by the theoretical frameworks that guide this study. For this thesis, the theory-driven codes were developed based on Lewis & Westlund's (2015) call for a sociotechnical emphasis to journalism research, as well as Guzman's (2018) human-machine-communication lens. It is important to emphasize that the theories informing this thesis do not lay out a deterministic view of technology as being the impetus driving drastic changes to the journalism industry. Instead, it aims to understand the diverse relationships between journalism and technology, and the extent to which journalists, programmers and technological tools and processes are becoming interconnected, and what this means for current and future newswork in Canada (Lewis & Westlund, 2015).

4.1. Self

The first code (Self), describes the interview participants' professional/personal identity and more precisely, how they define themselves as individuals and how they interpret their role and function as news media professionals within society. Although Lewis & Westlund's (2015) theoretical framework is concerned with acknowledging technology's distinct role in sociocultural research being done about journalism, and the inherent tensions between human and machine approaches—an aspect they believe to have long been neglected in the literature—themes of self and professional identity were addressed by several participants. This discussion about identity was found both in talk about personal experiences and beliefs, and in talk about how their work is fundamental to a functioning democracy, with little mention of the technology assisting them in their news production. Perhaps this can be explained by the fact that the past 20

years of journalism studies have turned towards trying to better understand the role of technology in newswork, and its implications for the social and organizational structures of journalism, while keeping human journalists at the forefront of the research (Lewis & Westlund, 2015). Lewis & Westlund (2015) argue that these efforts to understand technology's place in media organizations fall short, because this line of research pays more attention to human-centric considerations with little-to-no emphasis on the distinct role of technology. However, themes of self and identity that emerged from the data were considered important to discuss due to the participants use of automated journalism tools and processes, which in turn could provide insight into the ways these individuals give meaning to their newswork and how they identify themselves vis-à-vis the AI-enabled technologies they employ (Deuze, 2005).

Eight of the nine interview participants identified their roles as information providers, whose job it is to educate the public by providing context in a comprehensible manner. For Journalist 9, being a journalist means "trying to unpack and understand complex things in a way that is relatable to regular people." Journalist 5, the vice-president of his media organization's digital data desk, said he viewed his role as being that of a "professional storyteller," stating that "[journalists] try to uncover information that otherwise wouldn't be covered and that is of interest to people." Some participants addressed the moral obligation of newswork on an individual level. In one instance, Journalist 7 said the following: "I think being a journalist to me means reporting on the world, reporting facts and then interpreting it and disseminating it to an audience." Journalist 8, a working data journalist who also teaches journalism at the university level, sees himself as an "information professional who is there to tell society what's going on." Similarly, on the reason why he became a journalist, Journalist 6 said it was because of his interest in what goes on around the world, which allows you to "communicate something that is

newsworthy to your audience." The only freelancer of the group, Journalist 2, stated that "[people] don't get informed without journalism. So, I think there's my fundamental role of just, both transmitting information that is hard to get and for information that is easy to get, and to contextualize it." A journalist, according to Journalist 4, is "someone who likes to keep the public informed."

These shared views of the 'professional self' as a means of serving in the public interest and to a certain extent, keeping people educated, echo what journalism scholars like Deuze (2005) call a "shared occupational ideology" among news workers which serves as a way to self-legitimize their position in society. As Russo suggests (1998, in Deuze, 2005), "journalists identify themselves more easily with the profession of journalism than for example with the medium or media company that employs them." Indeed, the generally-shared sentiment of being information providers also turned five of the discussions with participants toward their social responsibility as journalists. Participants discussed their roles as serving a democratic function. Here, there was a leaning towards their social responsibility in educating audiences, who can then make informed decisions. One journalist explained it this way: "Fundamentally, I don't think asking citizens to make choices in a democracy works unless they're informed" (Journalist 2). Much like characterizing themselves as information providers, this view of journalists providing a public service is not new. According to Deuze (2005), "journalists share a sense of 'doing it for the people,' of working as some kind of representative watchdog of the status quo in the name of the people." For Journalist 1, being a journalist means providing news "in order to make decisions that are fully informed and to hold powers accountable for what they're doing." On the reasons given for becoming a journalist in the public interest, Journalist 2 said the following:

> I became a journalist because I think it's pretty vital to how our country works, to how democracy works. And you know, I was pretty disillusioned with the people who were running for office and the people that we elect. And I figured the way that I can contribute to making our country better is to do this. (Journalist 2)

Interestingly, only two participants (Journalist 1 and Journalist 5) brought up objectivity as key to the professional self-perception of journalists (Deuze, 2005). Striving for objectivity in journalism is not new, although it can be seen as a problematic way of thinking, as academics and journalists alike have revisited objectivity and admitted that it is almost impossible to maintain "value-neutrality" in newswork (Deuze, 2005; Pavlik, 2000). In relation to previous discussions about watchdog journalism and serving in the public interest, objectivity can be viewed as the overarching goal when promoting the field's democratic purpose. Journalist 1, a digital news editor and content strategist, had this to say:

> It's so important, especially nowadays, to inform the population in ways that are nonpartisan. I know a lot of media tends to be on one side or another but I really, truly believe that offering both sides of the meddle is where we get to inform people the best we can.

For Journalist 5, being objective is the overarching value that lends itself to the identity of a successful journalist:

> I'll say [being a journalist] is a commitment to being unbiased. It's a commitment to taking yourself out of the story and looking at it from all sides, not both sides, not a binary thing where I've talked to a crazy right-wing guy, now I gotta find a crazy left-wing person to counterbalance. It's more, what's the truth? What's the accuracy here? And if it's not a case of fact, then you present multiple points of view. It's not about you. If you want to keep up with you, then you can be a columnist, and there is a role for that in journalism.

Moreover, this non-partisanship ideal brings up interesting questions with regards to upholding journalistic standards. If journalists identify their professional-self as being able to disseminate objective news to an audience, how, if at all, do they continue to uphold these standards when research has already pointed to the inherently programmed intentions that are built into AI-

enabled technologies (O'Neil, 2016; Noble, 2018; Diakopoulos, 2016; Broussard; 2018)? This topic will be broached in greater detail later in the findings section.

Admittedly, this thesis assumed that the nine participants, whose expertise in automated journalism tools and processes sees them building the AI-enabled software, managing the technological products and/or overseeing their distribution throughout their newsroom, would not discuss their professional identity without acknowledging the distinct technological affordances that shape their journalistic output to a certain degree (Lewis & Westlund, 2015). Only two of the nine interviewees discussed technology's role in their professional-self perceptions. Journalist 3, a self-described data journalist had this to say: "There are some that are very excited about new technological tools, about what possibilities they open for not only storytelling, but news gathering and research analysis. I am one of those." In terms of using technology to achieve the "ideal-typical values in journalism ideology" (Deuze, 2005), Journalist 5 stated the following:

> [A journalist] is somebody who I think today has to be flexible, that takes part in what we're talking about. I think the idea of data and automation and using natural language processing and things like that are part of journalism. It's the willingness to find those tools, use them effectively to do the stuff we just talked about, about finding the truth, telling stories and uncovering stories.

It is important to note that the two comments above from Journalist 3 and Journalist 5 were considered of value despite not being a subject addressed by the other news media professionals. However, as Saldaña (2016) points out about the act of coding in thematic analysis, "frequency of occurrence is not necessarily an indicator of significance" (p. 41). For this reason, the discussions above were considered useful for this thesis, as it captures the essence of the object of study.

The sparse discussion representing technology's role in developing participants' professional-identities was particularly surprising in light of the fact that, as previously mentioned, algorithms and automated processes can now replace previous journalistic activities performed by humans (van Dalen, 2012). In revisiting Russo's (1998) view that journalists identify themselves more with the profession than the actual medium or media company, perhaps the data collected from these interviews adds to Russo's argument by including the dissociation between journalists and technology. If algorithms and automated journalism are to be viewed by journalists as what Woolley, Boyd & Broussard (2016) call a "civic prosthetic," then the human reporter must redefine his professional identity to encompass the sophisticated systems that assist or uncover stories that serve the public interest. Yet, Anderson (2013) and Lewis & Westlund (2015) offer an explanation to this professional detachment from technological affordances, suggesting that perhaps journalists still discount technically-enabled forms of journalism because of this shared belief that reporting, traditionally a human endeavor, is central to their professional craft and key to their agency.

4.2. Self in Relation to Technology

Self in relation to technology describes the relationship between journalists and the automated technologies they employ, particularly with respect to how they interact, the degree to which these tools and processes have changed their job and what is required of them, and their general considerations and observations on the extent to which they rely on technology to fulfill their specific journalistic tasks (Lewis & Westlund, 2015). This section of the findings offers a look into the varying levels of interconnectedness between news media professionals and automated journalism based on the insight offered by the nine participants who work at all levels

of the news production pipeline. From university-level educators and data journalists, to the managing editors and directors that oversee the distribution of these tools and products throughout the newsroom, the discussions with each individual provided an interesting view of the dynamic relationships between people and technology, particularly with regards to automated journalism in Canada.

Since all nine participants were recruited because they were known to be working with, or integrating, automated tools and processes in their journalistic practice to some capacity, it's no surprise that they generally held their working relationship with this type of technology in high regard. However, in light of the fact that not all participants have the same professional role/title, this portion of the results takes this difference into consideration, and begins by grouping the findings based on their professions to some extent. This approach made the most sense for this thesis based on Lewis & Westlund's (2015) holistic approach to understanding the interconnectedness between actors (news media professionals) and technological actants. Beyond simply focusing on the interplay between journalists and technology overall, these findings consider the fact that editorial workers are, as Lewis & Westlund (2015) suggest, negotiating "issues of authority, identity and expertise in connection with machine-led processes." Moreover, they do so in ways that differ from data journalists or university-level educators who, for example, while implementing automated journalism in their own reporting, are also tasked with educating journalism students so that they can eventually build, or work alongside, these automated processes.

4.2.1. Relationship with technology

As individuals with varying hierarchical positions within the journalism industry, there was an interesting difference in how participants viewed their relationship with technology, particularly in how they acquired these skills and their motivations for pursuing this line of work. For one, participants who described themselves as ***data journalists*** ascribed their computer-programming and automation skills to self-learning, and having completed university degrees in fields outside of journalism. The data journalist freelancer who was interviewed (Journalist 2), said: "I've been coding forever. [C] was my first language when I was 11. I pretty much used code in whatever I've done. I used to work in medical research and I did physics and mathematics in university." Similar pathways emerged among the other participants:

> I studied biochemistry and then computer programming, but I've been doing journalism since high school and CEGEP, like being involved in student papers. That's what interested me. So it seems like what I was studying was irrelevant to my current profession, but turns out it isn't. Now, everybody wants to do automated journalism and things like that. We all want to use technology to find and tell stories. (Journalist 6)

> Before getting into journalism, I got a bachelor's degree in physical geography. That's what I was interested in. What I wanted to do was cartography assisted by computer programming […] During my undergrad, I took every possible computer science class available. At the time, I wasn't at all a programmer or web-developer, and then I got into journalism and figured I'd never use those skills, because back in the 90's, journalism wasn't taught alongside computer science. (Journalist 8)

> So, I'm entirely self-taught. I've always kinda been that way. I think my preferred way of learning, is to just struggle through it alone. That's always been the case. I learned to code back in middle school when I was making GeoCities websites and whatnot. So that was a long time ago. Obviously, that wasn't like real coding like I do now, but it had all the same foundations, just getting used to working on a computer and learning how to make something appear on a webpage the way you want it to appear. All those things are kind of foundational. (Journalist 9)

The discussions with the participants mentioned above fall in line with the type of journalist Usher (2016) describes as "hacker journalist." Hacker journalists, as defined by Usher, are individuals who have developed their coding skills outside of the traditional journalism route (pp. 85-86). Instead of immediately pursuing journalism, they often focus on the development of

their coding skills first (p. 88). However, Usher makes a point of differentiating hacker journalists from programmer journalists. Programmer journalists, according to Usher, are usually self-taught and acquire their programming skills after establishing themselves as journalists (p. 86). If we consider Journalist 6, for example, who always had an interest in journalism despite never going to journalism school, it makes it difficult to place him in either the hacker or programmer category.

Yet, the following comments from Journalist 4 and Journalist 7 validate Usher's definition of programmer journalists, as they did go to journalism school prior to becoming proficient in programming languages. However, their decision to learn how to code was made to set themselves apart in the journalism market, emphasizing the highly competitive nature of the industry. One interviewee (Journalist 4), who is the head of his organization's digital data desk, mentioned teaching himself programming languages back when he was in high school:

> I started thinking, like, where is the growth area in journalism? And data journalism work seemed like that could really be the growth area […] So a job came up at [this media company] and they hired me. I convinced them that data journalism and automation were valuable, and I continued to hone my skills from there.

Similarly, one interviewee (Journalist 7), stated that he picked up data journalism because he was looking for a way to break into the Canadian journalism market. "It was really hard to find steady work. I graduated in 2009, but you know, we were in the middle of a recession […] so a ton of experienced journalists were out of work." That is when he started narrowing his journalism focus. As recounted during the interview, he asked himself: "what can I actually do that would differentiate me from the sea of other experienced people? And that's when I started doing data journalism."

Moreover, these findings suggest that perhaps journalists do not view themselves as falling into one camp over the other (hacker vs. programmer). Instead, they diverge from these

definitions and emphasize their interest in learning to code as something that gave them a specialization, with the added bonus of setting themselves apart with this coding "edge" as they navigate the journalism job market. Although this does not attest to the skills that may be required from humans in order to compete with the ever-increasing integration of automation in the newsroom, it does point to the pressures that may be felt by journalism students and journalists who are just starting out. The existing literature has already revealed that journalists are feeling an added pressure to compete with automated journalism software, which can produce news at a speed and scale that are unmatched by humans (van Dalen, 2012; Stray, 2016; Lokot & Diakopoulos, 2016). It is reasonable to consider then, that journalists may be increasingly learning to code, whether on their own time or in school, with a 'if you can't beat 'em, join 'em' mentality, in order to remain employable and competitive as interest for these newsroom technologies continues to grow.

4.2.2. Level to which technology has changed their job

Interestingly, individuals who discussed the ways automated journalism has changed their job the most were those at the management level. Journalist 1, a digital news editor and content strategist, said the following:

> The way it's changed my job is that I went to get a certification in product management just because I hadn't done any product management before. Now that I oversee the creation of AI products for the newsroom, I wanted to make sure that things were getting done properly.

Prior to the integration of automated journalism, Journalist 1 mentioned that he was the digital news editor. "So what I still do is assist reporters in telling their stories in a way that is clear for an online audience," he explained. "Now I need to ask myself, 'Ok, how do we make sure that we produce a tool that has a benefit for everyone in the newsroom?' And that wasn't a natural

step for me." For Journalist 1, acquiring new skills in AI product management and development was the most impactful change to his role with respect to his newsroom, acknowledging that automated tools and processes changed what was required of him. However, this shifting role is not only for reasons ascribed to self-interest, but rather as a means toward improving his knowledge and understanding of automated journalism so that the entire newsroom can benefit. This was a shared sentiment among others in managerial positions, whose jobs changed in terms of how they could help others, which in turn increased collaboration within their newsrooms:

> In some ways doing digital interactive, we kinda became the marginalized part of the newsroom, because not everybody needs to interact with you to get those interactives done. Now everyone wants to interact with us. The idea of automated journalism and using data to power all of that excites everybody. So all the different, you know, departments and desks and bureaus across the country are all really keen on what we do […] The possibilities of doing this are really limitless right now. We're just scratching the surface. So for me, my job, which is supposed to have innovation as part of it, got a lot cooler (Journalist 5)

> If I can automate 10 hours out of someone's job, they should be spending those 10 hours trying to come up with a new product or service that we can sell instead of just producing more content for the wire, let's say. [At my old job] I wrote some software that did [a former colleague's] work in about 30 seconds. (Journalist 4)

Other ways in which automated journalism was discussed as changing individual roles was at the local news level. Journalist 7 mentioned that it increased his ability to be hyper-local, to "hyper-target," stating that it allows him to "better serve [Canada's] news deserts, giving smaller communities access to the stories they want." According to Journalist 5, there is no newsroom in the country like his with "a dedicated data desk that blends data management, automation and human oversight," in order to provide a local focus. He had pitched this idea to a large tech organization that awarded a large sum of funding to four Canadian newsrooms, and that spawned what his newsroom calls their "digital data desk." Thanks to that external funding, Journalist 5 stated that they were able to help "the local news dearth that's in the country […] If we're talking

about how automation changed my job, that's one of my big focuses right now—getting that off the ground."

Attributing an increase in local coverage to algorithms is not new. In an article published by the *Columbia Journalism Review,* Diakopoulos (2018) writes about the Newsworthy tool developed at Journalism++.[2] It is software that does algorithmic claim spotting, which not only helps with the journalistic fact-checking process, but also with story discovery tools that can speed up the ability to surveil social media and large sets of data in order to uncover a newsworthy story or interesting leads. These algorithms "offer a sort of data-driven sixth sense to help orient journalistic attention" and can be adapted for different communities "by putting local aberrations in context with national trends, which helps reporters pursue local angles" (Diakopoulos, 2018). The similarities in the descriptions given by Journalist 5 and Journalist 7 about the scarcity of local news, and the way automation is developing the potential to cover these areas in the news cycle, is striking. Although this thesis considered the bulk of the literature on automated journalism, the advantages associated with automated journalism have been efficiency and productivity (Broussard, 2015; Stray, 2016), with quick turnaround times and low-marginal costs (Carlson, 2015; Montal & Reich, 2016; Lokot & Diakopoulos, 2016; Graefe, 2016). While less has been said about technology's ability to expand coverage specifically at the local level (with the exception of Diakopoulos, 2018), automated processes have been discussed with respect to generating news articles that may not have been done in the past due to limited time and resources. While discussing how his role changed to incorporate overseeing the

[2] Journalism++ is a network of international data journalism specialists that was founded in Sweden in 2011. They integrate sophisticated data analysis, data-driven storytelling, programming and full-stack design to produce journalistic work.

production of "locally relevant news," Journalist 5 gave an example of how an algorithm would be used for this purpose by his newsroom:

> The ability to take a national data set—and it could be crime, it could be health outcomes, it could be economic indicators—and make it locally relevant, automating a story so that we're going to put out say, the Brandon, Manitoba story or the Corner Brook, Newfoundland story automatically. I think that's huge for us. So, that will help sort of fill up the, you know, the bucket of local news that's really been emptied over the last probably 10-15 years, as many of our [wire service] users have downsized.[3] Papers and news outlets have shuttered. This is a way to keep people locally having that information. So, I think that's really good.

Moreover, Diakopoulos (2018) suggests that algorithms can "turn a statistical lead into an interesting news package with local context and sourcing." It is worth pointing out that Journalist 5 attributed this new facet of his job description to receiving funding from an external source. While he is hopeful that local journalism could bounce back with the help of algorithms, he acknowledges that media organizations have been downsizing or shutting down because the money simply isn't there. Furthermore, these findings suggest that perhaps the possibility of expanding local coverage is limited to those who have the resources to achieve this feat, making his news organization an exception within Canada's current media landscape. Despite the recruitment for this research only amassing nine news media professionals, no participant from any other news outlet spoke of having the necessary funds to expand their ability to cover local journalism, thus making Journalist 5's comments relevant to this research, as they point to the need to revaluate journalistic business models—a theme which will be addressed later in this chapter.

Journalistic collaboration and hyper-local journalism are examples that fall within Lewis & Westlund's (2015) characterization of "cross-media newswork," which refers to the integration

[3] Although participant names and their affiliate media organization have been omitted from this thesis for ethical purposes, it is worth mentioning that Journalist 7 works for a media organization that provides newsrooms with a newswire service. When he refers to "users," he is referring to the news outlets that pay for the stories coming from the newswire.

of multiple media platforms and acknowledges the various forms of journalism produced within this holistic, sociotechnical framework, including "editorial as well as business and technology activities, thus rendering a more complete picture of news publishing at the organizational level." Journalist 3, a data journalist working at a large Canadian media organization, explained it this way: "I think it's more that it created a specialization, right? More newsrooms are seeing the value of hiring dedicated computational journalists or data journalists who specialize in this tech." Journalist 3 went on to say the following:

> There are specializations in newsrooms these days that never existed before, like the journalists who can code. This is a very, very new phenomenon. The journalists who can be a liaison between the editorial and the development desk […] So there is more and more of this melding of storytelling with computer science.

Other interview participants who did not have editorial or managerial roles per say, discussed automated journalism as not really changing their jobs, but more as an opportunity to enhance their reporting, or in teaching journalism students to enhance theirs. "One of my role models is Philip Meyer,[4] who is now at UNC, and in one of his books, he starts by saying: 'I don't know if you've noticed, but they've upped the ante on what it takes to become a journalist,'" said Journalist 8. Knowing how to code and apply AI-enabled processes to your reporting, according to Journalist 8, "is an added skill that journalists of today need to have […] You become better able to tackle any questions that are thrown your way, because you learn how to work with data to find answers."

[4] Philip Meyer is a journalism professor and former holder of the Knight Chair in Journalism at the University of North Carolina at Chapel Hill. He is considered to be one of the earliest adopters of computer-assisted reporting (CAR), after he analyzed survey research on a mainframe computer during the riots in Detroit in the late 1960s. He was able to show that college-educated people were just as likely to take part in the protests than high school dropouts. In 1973, he authored *Precision Journalism: A Reporter's Introduction to Social Science Methods*, which is currently in its fourth edition.

Journalist 9, who was initially hired by his media organization as a graphics editor, said that his ability to automate tasks and processes changed his job because he managed to make the position he was hired for obsolete:

> It definitely changed my job in the beginning, when I was hired for one task and I automated my job out of existence within about a year and a half. It was only by virtue of building that first charting application that I was able to do all the other stuff I wanted to do, because I didn't have to do that first task anymore. It was only taking me 10 minutes a day as opposed to six hours a day.

As for Journalist 2, his role remains unchanged: "Pretty much everything I've done in journalism has been combining my knowledge of technology and machine learning with journalism to create original, cross-platform multimedia journalism," although he does point out that it "simplifies some tasks of journalism" in a variety of ways, such as uncovering stories:

> Another is to identify patterns and find stories that would be too labor intensive to find by hand. I think, you know, we've seen some stories—like there was one in Ukraine where they used satellite imagery and machine learning to identify illegal amber mines. That sort of thing would be by hand, to analyze all of this satellite imagery, impossible. It would take a hundred years, you know. So, you get new stories.

Similarly, Journalist 3 mentioned that automation "makes my job easier, [allows me] to cover more, to have a bigger antenna for things that are happening, and saves time from having to manually go to different places on the web, to call different people." He mentioned that the information is "coming to [him] instead, in one big sort of like firehose of information that [he] can then filter."

Many of the points of discussion mentioned above fall in line with what has been said in the literature with regards to automation serving as a time-saving tool by taking over tedious tasks (Carlson, 2015; Broussard, 2015; Graefe, 2016; Lokot & Diakopoulos, 2016; Dörr, 2016; Montal & Reich, 2016), particularly with regards to boosting productivity and increasing workflow. The sub-theme that follows (Level of reciprocity) discusses how, to what extent, and

for what purpose each participant uses automated journalism in their daily newswork. Yet, the findings presented above demonstrate that the overwhelming majority of interview participants view the opportunity to apply these sophisticated technologies to their newswork as a way to increase newsroom collaboration and to expand coverage. Albeit, they do so while also acknowledging the rising pressure to have a specialization in this field of journalism.

4.2.3. Level of reciprocity

The "level of reciprocity" sub-theme provides a comprehensive overview of the ways in which each participant discussed exercising automated journalism tools and processes in their news production, the extent to which these AI-enabled methods are employed, and what functions they serve. In exploring automated journalism using Lewis and Westlund's (2015) theoretical framework, these technological tools and processes are referred to as technological actants, which they define as material objects "that are notable for their association with human actors and the activities they undertake in conjunction which such objects." As aforementioned in Chapter Two, Lewis & Westlund (2015) borrow from Latour's Actor Network Theory (ANT) in tracing associations between human actors and actants. However, ANT does not distinguish between human and nonhuman actors, nor does it privilege one over the other when studying the interconnectedness of these two entities. For this reason, they take Latour's query (2005, p. 71, in Lewis & Westlund, 2015), which notes "the questions to ask about any [actant] are simply the following: Does it make a difference in the course of some agent's action or not?" and modify it to highlight their sociotechnical approach. The question then becomes: "Does a technological object like cross-media news work, application programming interface (API), or a set of software code make a difference in the course of some actors' activities or not?" (Lewis & Westlund,

2015). This line of thinking provides a new way of assessing the ways in which technological actants are created and instructed by humans to suit journalistic purposes in the production of news.

Several participants brought up using machine-enabled tools and processes to ease their workflow, offering an answer to Lewis & Westlund's (2015) question about whether or not technological actants make a difference in the daily activities of human actors. Yet, for some, adopting these technologies shouldn't be seen as an option with the upside of offloading some aspects of journalistic work to various actants. Instead, it was discussed as being a requisite:

> A big part of my job is innovation and trying to push it forward in the newsroom. I can't say it's easy, but I think it's a requisite. I think it's necessary for newsrooms nowadays to adopt technologies that may ease their workflow. And not just ease their workflow, but I think the request that's being made out of journalism now is different than what it used to be. You know, we rarely have three months to work on an investigation. We need tech to come and help us out so that we can push these things further and modernizing the workflow in the newsroom I think is where [technology can] assist us. (Journalist 1)

For Journalist 4, his primary reason for automating certain tasks is to help his colleagues in the newsroom: "They were doing a lot of stuff that just made no sense to me, like how long it took them to do these things when they can be done much, much faster [with automation]." He then provided an example of a former colleague who would do the sports aggregate for her newsroom, where she would have to go to every different minor sports leagues' website in her province to retrieve game statistics—a feat that would take her hours to accomplish every time. "I wrote her some software that did it in about 30 seconds. And it was just, you know—now she can spend her time doing something she enjoys. She can do actual reporting, she can do whatever she wants, to take the time to learn a new skill."

Many interview participants mentioned that automated journalism has helped ease journalists' workflow by automating tasks that are considered to not be the most effective use of

human journalists' time. "It's about simplifying some of the tasks of journalism […] that people hate and freeing up journalists to do things that are a better use of their time" (Journalist 2). This idea was complemented by the fact that participants believed some of these tasks to not be what reporters consider journalistic work at all:

> I figure if a computer can do it better than a human, why don't we let the computer do that and let the human do something that only humans can do. At my [current media organization], I've really been pushing the same idea. Like, I don't think people get into journalism to do the things that they end up doing […] I automate stuff so that they can actually focus on other things. (Journalist 4)

Prior to the integration of automated journalism in his newsroom, Journalist 5 said that some of his journalists were essentially doing what "robots do," such as rounding up the scores of different sports games and publishing game recaps, or content optimization (i.e., rewriting headlines, keywords). He went on to say that "nobody went to journalism school to do that. It's literally a robotic function and it's boring and mind-numbing, but it's important because our clients want it." For Journalist 9, he considers the things he has automated in his newsroom to be "low value, grunt work type stuff—things you would have to do by hand." By automating certain tasks, he was able to free up a colleague who is in charge of the business beat. "He can focus more on asking more detailed questions," for example, whether it be "about job growth starting to slack in any given city, or what's happening to the GDP of a province. Again, you're trying to automate out the low value stuff." Journalist 8 maintained that "automated journalism can answer, can help to fill in some of the tedious tasks in journalism. The best example is the *LA Times'* Quakebot."

It is important to underscore that what is perceived by the participants as "tedious tasks" or "low value, grunt work type stuff," were discussed in relation to what clients (news outlets) and audiences wish to consume in terms of news. In providing Quakebot as an example of

automation taking over aspects of journalistic work to generate news output, Journalist 8 highlights the fact that even if monitoring the U.S. Geological Survey's Earthquake Notification Service for an earthquake alert is considered to be a tiresome task, it is still a piece of news that is invaluable to the public and essential to disseminate. By replacing a human who was tasked with surveilling the U.S.' earthquake monitor, the *LA Times* is still serving in the public interest, only this time with speed and scale, which is the overarching goal of the integration of automated journalism in the newsroom (Carlson, 2015; Graefe, 2016; Diakopoulos, 2019).

In order to better understand how these technological actants are easing the journalistic workflow, the following provides detailed examples given by the nine news media professionals and are subsequently categorized based on the types of tasks they help accomplish. This is then followed by the types of stories produced (i.e., sports, business earnings, crime reporting).

Types of tasks being accomplished

Web scraping and data mining

Web scraping and data mining were brought up by several participants as an automated process that they regularly implement, and as a tool that contributes to their newswork. Web scraping is an automated process used to gather (or harvest) publicly accessible/available data from various websites, thus allowing an individual to extract data that is of interest for further analysis and use. Although data mining serves a similar function, it involves discovering patterns in large stores of data that are not necessarily on the Web, and is made possible through the use of sophisticated statistical algorithms. Journalist 7 highlighted the importance of differentiating between Web scraping and data mining. To him, the latter has more of an impact on journalists:

"[Web scraping] is just the basic retrieval of information and that's been going on in newsrooms for decades. Using algorithms to actually create something, like removing a significant part of the journalists' task of synthesizing information, is huge." Journalist 9 also reiterated the fact that Web scraping has "always kind of existed in newsrooms." He added that it's "not really that this is anything new, it's just a constant evolution. We're constantly looking to do more with less."

Data mining, according to Diakopoulos (2019), "is often enabled by machine learning techniques—algorithms that allow for computers to learn patterns from the data (p. 44). Ultimately, both processes automate "research and news gathering" (Journalist 3), which are enabled by running "statistical programming languages and more advanced approaches to crunching datasets for storytelling purposes" (Journalist 9). Journalist 6, a Canadian interactive graphic journalist now working in the U.S., stated that "most [media companies] like ours will have a scraping product to scrape any website and to be able to watch those websites, especially in the context of financial news." This form of data scraping enables journalists to access new information with speed. "You know, if something comes out on a [financial] website, you want to be able to see it immediately, within 30 seconds or less," emphasized Journalist 6. Journalist 9 mentioned that he built a tool for his newsroom that uses machine learning to "cluster a very large dataset to find interesting outliers." He provided an example of a tool that is currently in use by his newsroom to monitor social media: "We have a Facebook political ad collector […] which uses machine learning to determine whether an ad that we scraped from Facebook is political or not, and then it shows that to reporters for analysis."

Although scraping is a process that can be fully automated, it requires the individual who builds it to be proficient in programming languages, such as "R, Python, JavaScript and CSS" (Journalist 8) in order to write a script that can then automate the "retrieval of data from certain

sites" (Journalist 7). Yet, as told by Journalist 9, the result of the analysis is then shared with a reporter, thus demonstrating that human input is not merely reserved for the initial step of writing the script, but also at the end of the process when analyses must be rendered. If web scraping and data mining are to be considered as part of the frontiers of what is possible to accomplish with automation, Diakopoulos (2019) argues that human journalists still have a lot to add to these types of tasks when it comes to "complex communication, expert thinking, and ethical judgement," which in turn increase the hybridization of algorithmic and editorial thinking (p. 40).

It is worth mentioning that in order to build or work with automated journalism tools and methods, journalists must be skilled in algorithmic, social scientific and statistical newswork (Anderson, 2012; Diakopoulos, 2015), which is why Journalist 8 teaches programming in his data journalism courses: "All of our students must learn programming languages […] and that's when it gets a little more complex. But why do they learn? It's so that they can use scrapers and apply natural language processing." Data journalism is a term that Usher (2016) believes to be "quite slippery" as it can mean different things to different practitioners (p. 22). Yet, the clearest definition she relies upon is the one given by Journalist Alexander Howard (in Usher, 2016) in a report for the Columbia School of Journalism: "[The] gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism" (p. 22). However, Usher points out that not all data journalism is interactive, but that the form of data journalism rendered in interactives is the one that relies on a technical skillset that extends code into journalism, much like the form exercised by some of the interview participants (p. 22). During the interview, Journalist 8 stressed that "the goal of teaching these methods to [our students] is

not to turn them into programmers, but to turn them into people who can demystify algorithms […] and understand how they are built."

It's no surprise that web scraping and data mining were the most predominant methods of automated journalism addressed during the interviews. As Graefe (2016) points out, "the most important driver of automated journalism is the increasing availability of structured and machine-readable data provided by organizations, sensors, or the general public." On the other hand, Anderson (2012) attributes data-driven reporting to open-government initiatives:

> The role of large-scale databases in the crafting of public policy, the transparency and accessibility of government data and its use in democratic decision-making, government surveillance of online data and political activities: all these state-level initiatives may ultimately influence the forms of computational, data-driven journalism that news organizations adopt. (p. 1011)

As will be discussed later in this section in the outline of the types of stories produced, data mining has six distinct capabilities: classification, regression, clustering, summarization, dependency modeling, and change and deviation detection (Diakopoulos, 2019, p. 46). These capabilities help journalists detect and discover what's in the data and what is considered newsworthy, which ultimately depends on what dimension of a story a journalist is going after (p. 47). Moreover, the adoption of automated methods for data aggregation and analysis make sense based on the increasing availability of large sets of data from all levels of society. It provides a better understanding about the manner in which newsrooms and individual journalism actors integrate and work with automated technological actants, and how they become embedded into their workflow.

Monitoring and tracking changes

Using automation to monitor and track changes was brought up by the participants as another way in which automated journalism facilitates their workflow. Two participants (Journalist 3 and Journalist 9) mentioned using the open-source reporting software Klaxon, developed by the Marshall Project.[5] It enables reporters to monitor websites of their own choosing, including government websites for newsworthy changes. This web-based program is aimed at journalists who know how to code, and who want to set it up in their newsroom and potentially contribute to Klaxon's continual development by applying their knowledge of programming languages, particularly Ruby on Rails. "Klaxon watches websites for changes every 10 minutes, and if there's a change it emails me or other reporters depending on the websites that we decided to watch," said Journalist 9. "[It's] quite important because in a reporter's daily job, there's a million things you're trying to keep track of and it's very easy to be inundated by Google alerts or you know, have 47 tabs open where you're checking every single one every day" (Journalist 9). This semi-automated tool, according to Journalist 9, removes some of what he called the "low value grunt work from our daily list of tasks." It is especially useful for monitoring government activity, as described by Journalist 3:

> What it does is you basically, say, give it a website to monitor and every time that website changes, it alerts you, right? This is good for seeing when the government changes its position on something. Maybe there's a controversial issue that's going on right now. I didn't use this for it, but recently the Quebec government made an immigration reform. Maybe during the elections, you can monitor a Party website for platform changes if they changed their website without telling anyone. They can do it on the sly, but [Klaxon] can let you know about it. Let's say, specifically I ask the RCMP to give me the complete database of missing people and unidentified bodies, which is online. You can search through it, but you can only search say, 10 profiles at a time. You have to keep clicking on next, next, next. I have to click on the profile to see the things. Well, what if I want everything on my machine so I can see, okay, when are people disappearing? What kind of people are disappearing? Do they have a certain look? A certain age? The government won't give you that […] You harvest that information and get a computer to do what a human would do, which is to go

[5] The Marshall Project is a non-profit news organization that focuses its reporting around the U.S. criminal justice system. They are primarily known for their collaborative investigations and data reporting.

> through each one of those profiles and download it to your machine. It does it in minutes instead of having to click for hours.

"Big investigations are definitely an obvious application of this type of monitoring software," remarked Journalist 3. "Like for daily breaking news, maybe not so much […] but for the most part you have long term investigations that are very propitious to this kind of tech." Tools for monitoring and tracking changes can also be developed in-house if a newsroom has the skilled programmers to build them:

> I have a colleague who has written code in a statistical programming language called R, that makes it very easy for him to process incoming jobs reports and GDP numbers and whatnot. So when [Statistics Canada] posts new figures every month or every quarter, he can automatically pull down that data and it'll say, Okay, here are the interesting outliers this month. (Journalist 9)

These monitoring tools can also be applied to social media platforms. Journalist 8 built a Twitter bot that is linked to the Canadian Legal Information Institute's (CanLII) RSS feeds. His Twitter bot tweets as soon as court judgements from his respective province go up on the CanLII website. During the interview, he checked the last time his bot tweeted: "Its last tweet was six hours ago. It tweeted a new ruling by the [Province's] Court of Appeal." In discussing the advantage of this type of bot, Journalist 8 asserted that, "for reporters covering the courthouse beat, instead of having to check CanLII or subscribing to its RSS feed, if they're following this Twitter account, well, boom. They see the court decisions in real time." Another example of monitoring via Twitter was provided by Journalist 3, who built a bot to monitor every time a federal candidate retweeted during the 2019 Canadian federal election:

> I wanted to see if they were retweeting fake news or falling for misinformation, things like that. Imagine having to manually monitor a thousand Twitter accounts every day? One person can't do that. It's impossible. So I get my machine to monitor these Twitter accounts every time [a candidate] retweets, save it, run it through a service that checks if the account being retweeted is bot-like, and then you have a nice filter that's like 'Oh, these are the suspicious retweets that I have to pay more attention to. I wouldn't be able to do that otherwise.

These automatic or semi-automatic bots fall in line with what Lokot & Diakopoulos (2016) define as news bots, which are "automated accounts that participate in news and information dissemination on social networking platforms." Although the examples provided above outline how they can be applied as website monitoring and tracking tools, particularly government websites, Journalist 3 does bring up the fact that bots can be purveyors of fake news in his recounting of his Twitter bot that monitors candidate retweets. This was the first mention of any sort of algorithmic accountability reporting by a participant interviewed for this thesis (Diakopoulos, 2019), a concept which will be discussed later in this chapter. News bots, however, were highlighted by the participants as being one of the most prominent ways they go about accomplishing their journalistic work, going back to the idea of bots serving as a "civic prosthetic" (Woolley, Boyd & Broussard, 2016). In other words, these bots are a way to supercharge their reporting practices by taking over tasks that would otherwise take them hours to achieve without some form of automation (Woolley, Boyd & Broussard, 2016).

News bots can also directly interact with users and tailor to specific audience needs (Lokot & Diakopoulos, 2016). An example of this was given by Journalist 4, who developed a bot API—a set of functions (codes) that enables data transmission between multiple software intermediaries—for the last federal election:

> We brought in all the census data from the last census and we combined that with all the election data that we have because we do all the elections. I built this tool that we ended up selling to five or six [news outlets] that could run in Slack. So, you would go into Slack and it was built to be very conversational. It was like you talked to the bot and said hi, it would respond Hi, how can I help you? And give you a list of all the stuff it can do. And so you would say, okay, riding one, zero, zero, one, and it would bring up all the information about Avalon [Region in Newfoundland]. It would give you the population, demographics, education statistics, or whatever all the way through. And then you could actually drill down deep into the data. So you'd say population details, Avalon, and it would show you the very granular population, like 18 to 64, but then 18 to 21, 21 to 24 […] that sort of thing […] If

> you wanted to find out how many people in Avalon have master's degrees, it could tell you, and it was drilled down by riding detail.

This news bot, which functioned through the business communication platform known as Slack, had other impressive capabilities. "We offered stuff that you couldn't get just from the data. So it would also do comparisons. So for everything, you could see the rank, which is not something that Statistics Canada provides you," explained Journalist 4. He went on to say that it could also compare two ridings of your choosing, and even generate a 700-800-word profile for every riding. The user would simply need to type "Story, [specific riding]" and it would automatically generate a profile of that riding using natural language processing software. Although this bot was developed to interact with journalists from news outlets that bought Journalist 4's company's newswire service, he also spoke about another Slack bot built for radio broadcasters that "would give [broadcasters] the latest [election] results up to the second." In addition, he manages Twitter bots for his newsroom that automatically tweet out results and interact with their audience. "If people Tweet a reply asking for a map, it will respond with a link to an updated map." Based on these conversations, social media and monitoring tools are being embedded into the workflow of a growing number of reporters, and can have value for journalists, particularly with respect to the time-saving abilities of these automated methods in data aggregation, analysis and uncovering newsworthy stories.

In-house tools and applications

Although the section above mentions certain scraping tools and bots that were designed in-house, this section outlines other custom tools that were brought up during the interviews. Interviewees explained how their newsrooms, or they themselves, built custom search engines and database research tools to manage all of the data they scrape from websites or mine from

large data stores. In his newsroom, Journalist 1 mentioned that their automated research bots store information in "databases that we either own or that we can find through various entities, but that yield results in an automated fashion for us," which can be accessed by all of their reporters. Similarly, Journalist 6 revealed that while he was employed by a Quebec news organization, they were working on a story about donors to Justin Trudeau's campaign back in 2013, before he became prime minister. "The Elections Canada contributions database makes it really hard to find the names of donors […] It was very bad and we had to go through a lot of loops," he recounted. Instead, they downloaded the database: "We downloaded the database ourselves and we could do all of the tasks [we wanted] based on our [custom search criteria]. We basically built our own search engine." In that same newsroom, they had another tool that aggregated information about people who are in the public eye, namely politicians, prominent businesspeople and Canadian celebrities:

> We had a tool where we had information about different people. I mean, not in public databases, but we stored [the data] all in the same place. I don't know if you can call it a scraper, or bot or whatever. They are programs, I would say in general. They are programs you run in a regular fashion. It could be automated, could be manual. And then you collect information and then you put it in [your newsroom's] data store and then you use it for reporting. As a journalist in the newsroom, I think you can find use for that type of information.

Other instances where tools and applications were developed in-house concerned charting applications. Journalist 9 was initially hired as a graphics editor to make charts and visualizations for his media organization's newspaper and for the web. "A single chart would have taken 45 minutes to do by hand," he explained, which led to the creation of the charting tool:

> When we built this tool, the idea was to automate this process such that it now took less than 60 seconds and democratized that kind of work in the newsroom. So a reporter, if they were so inclined, could use the tool and build their own charts for their own stories, they wouldn't have to go through three or four different people to get that done.

Similarly, Journalist 6, an interactive graphic journalist, spoke of his process in creating a tool that could generate graphs:

> I think for graphics, our things are more bespoke. But despite this, even though it's bespoke, we can still automate a lot of the parts of those graphics that we make and make the job of updating them more easily. Like I see my job of building graphics as kind of, like you're trying to build a shell that's gonna fill up with data and you try to build it as generically as possible […] So you need to build these skeletons that can take generic data and then you fill them up with code, using code to make them beautiful, like colourful or whatever, making the shapes the right size.

While Journalist 6 discussed generating charts from a design perspective, the integration of automated design and publication tools were also imparted by some of the participants: "There are automated tools […] for publishing on the web. These are tools that just make it easier to put things up on the web that any journalist can do himself without having to go through several editorial steps," as Journalist 3 explained. These tools, he continued, also automate things like "embedding content […] they automate tagging the story with tags that make it easier to categorize so [journalists] don't have to do it themselves." Journalist 4 added that when journalists in his newsroom want to tweet an image, "it was always a hassle for people to put watermarks on the proprietary images." To solve this issue, Journalist 4 said that he built an "automated thing where you drop your image in and it sends it back to you with a watermark, and then nobody has to use Photoshop or Gimp. [It's] a web-based thing that you can do right on your phone." He also built a tool that allows journalists to use "cards when [they're] quoting someone for Twitter so they look nice," explaining that "[journalists can] upload them to Twitter and they're perfectly sized, and [we noticed] that we get more engagements on our tweets that way." "Basically, they just have to drop the text into a text box and it'll give them like 15 options of how they want it to look. They choose the one they want, download it, and they're ready to go," Journalist 4 added.

All of the interviewees acknowledged that AI-enabled processes have the ability to automate routine and repetitive tasks, which in turn can free up reporters to pursue more interesting forms of work, such as more in-depth investigations. Yet, Graefe (2016) points out that the extent to which "automated technology will replace or complement human journalists will depend on the task and the skills of the journalist." Journalist 2 mentioned that one prominent example of a journalistic task that is increasingly being replaced by automation is that of interview transcription: "I use auto-transcription. [Transcribing interviews] is something that I hate doing and it frees me up to do the things I'd rather be doing." It can be assumed that the act of transcribing interviews falls in line with the types of low-value, routine and repetitive tasks that were brought up by other interviewees, as being processes that "nobody went to journalism school [for]" (Journalist 5). Ultimately, the automation of this step of the news production process does not replace journalists, but instead exemplifies the complementarity that can occur between journalists and technology. In this sense, journalists benefit from having this type of task removed from their daily workload (Graefe, 2016; Woolley, Boyd & Broussard, 2016).

In contrast, Journalist 1 mentioned an automated tool that does have the potential to shift journalistic roles. During the interview, he explained that his newsroom has a big translation desk, disclosing that "50 to 70% of our French newsroom is tasked with taking stories from [our English branch] and flipping them to French." His newsroom built a tool that automates the translation of articles from French to English or vice versa. Journalist 1 explained why they chose to adopt it:

> The whole process [of translation] is kind of long if you have a four-page story. [Our translators] print it out, put it next to them and start translating manually. So, I developed a product with my team and with Google so that the story gets automatically translated.

It can be argued that the introduction of this type of automated translation tool has the potential to displace the people who were hired to fulfill the translations to some degree. However, shifting or evolving work processes are not new to the journalism industry (Dörr, 2016) and often "force new tasks on reporters and editors alike" (Powers, 2012, p. 27 in Dörr, 2016). Such is the case for this translation tool, as outlined by Journalist 1:

> With automated translation, we're saving a huge amount of time for every story. Is it perfect when it spits out a translation? No. So yeah, [our staff] needs to spend time editing it. We're not trying to [replace our staff] and save time 100%. What we want is to assist reporters in the way they're doing things […] Basically, our translation tool has filters added over top so that we can include all of our [newsroom's] terminology, like a dictionary. It then shows where those words are in the text and allows our editors to flip them and do more of [our media]-style edit to it. So we're slowly in the process of changing the translator job to an editing job […] The bulk of the translation is done automatically and then the revision part is where [our translator's] work is now crucial.

Based on the discussion with Journalist 1, the adoption of this automated technology did not replace anyone at their translation desk, but rather shifted their roles to something with more of an editorial oversight. As he mentioned, the translations are still not completely accurate, thus emphasizing that this task, like many others, still require human input, which can be attributed to the fact that humans have skillsets that machines do not, such as analytical thinking, creativity and a certain understanding of a newsroom's publication style (van Dalen, 2012; Stray, 2016; Graefe, 2016). Moreover, given the fact that Canada is a bilingual nation, it would make sense for Canadian newsrooms to want to adopt this type of tool. Although Journalist 1 discussed his newsroom's automated translation software as something that can facilitate journalistic workflow, while also personalizing the format of a story to fit their newsroom's style, Journalist 5 ascribed automated translation tools to the opportunity to expand coverage:

> If there's a story that has universal appeal, we want to translate either from French to English or vice versa, and I believe automation can help […] But I think the idea of, you know, look at Google Translate or some of the other online translation tools. They're getting better every

> day. I think that goes back to how the improvement of AI technology has been, I think, exponential over the last five years. It's been remarkable.

Moreover, the interview participants provided thoughtful examples of the many ways in which they utilize automated journalism tools and methods to assist them in their reporting or editorial tasks. Data mining, Web scraping, website and social media monitoring, chart building and page design are just a few examples of the ways in which automation can assist news media professionals in their day-to-day tasks. All of the participants spoke of these automation methods with excitement, which is reasonable given their ease with building or applying such technologies for the benefit of their newswork. Later in this chapter, this thesis will explore their individual perceptions about how their peers and more traditional media might perceive the expansion of automated journalism in newsrooms, and what this means for journalistic practice and standards. The extent and the speed with which the development of automated journalism will take place remains disputed (van Dalen, 2012), and particularly understudied in Canada. However, van Dalen (2012) remarks that "optimistic voices" argue that the application of algorithms and automated methods might change some aspects of the journalistic profession, but "they won't replace journalists just as spell-checking programs haven't replaced secretaries" (Bunz, 2010, in van Dalen, 2012).

Types of stories being produced

The literature acknowledges that the news media industry is among those that are carving out spaces for AI-enabled technologies, while simultaneously weighing the opportunities and challenges of algorithms and automation for journalism (Lewis & Westlund, 2015; Broussard, 2018; Diakopoulos, 2019; Lewis, Guzman & Schmidt, 2019). As mentioned in the previous section, the types of tasks being accomplished by these technologies can ultimately assist

newsrooms in content generation. During this thesis' interview process, participants revealed the various types of semi- and fully-automated stories they and their newsrooms are producing. Unsurprisingly, these stories centre around topics that are rich in structured data, namely sports, financial news, politics and crime. As the literature has shown, automated journalism and the sub-field of algorithmic journalism, relies on the availability of machine-readable data to coordinate computational inputs and outputs in order to make inferences (Carlson, 2015; Graefe, 2016; Linden, 2017; Diakopoulos, 2019). This portion of the findings looks at the examples of stories produced by the participants, and details how exactly they achieve the production of these stories.

Sports

In 2012, van Dalen published a paper about newsroom algorithms and the technical strengths that render them competitive with human reporters. He states that "the strengths of automated content generation are the low marginal costs, the speed with which articles can be written and the broad spectrum of sport events which can be covered" (van Dalen, 2012). Although automated journalism is no longer limited to sports coverage, it is a news beat that is increasingly relinquished to automated technologies in Canada and elsewhere. In Journalist 4's newsroom, they have automated all of the Canadian Hockey League (CHL) game recaps, which encompasses the three major junior hockey leagues in Canada, representing a total of 60 teams. According to the CHL website, more than nine million fans attended CHL games in the 2019-2020 season, which demonstrates a clear interest for this type of sports coverage. During the interview, Journalist 4 stated that the hockey games were his newsroom's first attempt at automatic content generation using natural language processing. He also explained why his newsroom chose to adopt this method:

> On any given night, there could be 30 CHL games. I don't think it's possible for one person to write 30 stories in one night. It just doesn't seem feasible. So then, how do you decide which teams get covered, and which don't? That's an editorial decision to make. By operating this [automatic content generation], all of them get covered. So now every audience gets something about their hockey team as opposed to just 15 of them. It allowed us to cover games that we wouldn't necessarily cover.

In returning to the notion that automation can save journalists from doing repetitive tasks, Journalist 4 reiterated: "You know, I don't think anyone who went to J-school was like, 'Man, when I graduate, I need to go work somewhere where I can write 200-word stories off of box scores.'" The automated hockey recaps were introduced because his newsroom realized that their clients, whether it be media organizations from all over the country or direct consumers, wanted to know the game scores for their home team. He then proceeded to outline how the automated process works:

> So basically, the [journalist] who works nights or whoever's doing it, goes to the Web app that I've built. They put in tonight's games and they click on the generate button. They have a few different options for what they want to put in the stories. And then they just copy it out and put it into the content management system […] The meat and potatoes of the whole thing has been automated for a few years now.

Although his newsroom has not yet applied this software to other sports stories, Journalist 4 mentioned that they do something similar when they cover the Olympics, using the same template generation process. In relation to sports coverage, Journalist 5 said that his newsroom has "talked about other sports possibilities," although he did not want to go into further detail.

Another interesting finding with regards to sports coverage was the fact that Journalist 1's newsroom, which operates as both a French and English newswire service, limits its sports coverage to English. "[Automation] has been mostly happening in English right now. We haven't tested [automated hockey recaps] in French yet. We're still holding off," conceded Journalist 1.

Finally, when Journalist 4 described the tool he built to automate the production of CHL game recaps, he noted that this was the only portion of news that his media organization had fully automated and launched to date, despite pilot tests for other projects being underway. Although evidence to support the following theory is limited, this could potentially indicate that newsrooms, like the one employing Journalist 4, might be reluctant to pursue automation for more complex stories. Graefe (2016) provides insights in this regard by detailing a 2015 German study that randomly assigned participants with sports and finance articles, which were either human-written or machine-written. The results found that participants perceived the automated and human-written articles to be of similar quality. "Although human-written articles were perceived as somewhat more readable than automated ones," Graefe writes, "people did not particularly enjoy either of them" (2016). According to Graefe, the general dissatisfaction with both stories might pertain more to the content than the writing: "[Sports and finance topics] are routine and repetitive tasks […] and comes down to a simple recitation of facts and lacks sophisticated storytelling and narrative." Moreover, sports are rich with statistics and machine-readable data, which make it suitable for automatic content generation. The other types of stories that were mentioned by the nine interview participants and further outlined below, provide examples of semi-automated news production.

Financial News

Financial news, much like sports, is a topic that has been widely associated with journalism's turn towards automation, since it generally pertains to quantifiable information with "good input data availability" (Dörr, 2016). Automatic content generation companies, like Narrative Science and Automated Insights (See Chapter 1), initially focused their automated

writing algorithms on journalistic products well-suited for sports and finance (Diakopoulos, 2016; Graefe, 2016; Montal & Reich, 2017), because according to Dörr (2016), these domains consist of "processable and structured data that are available in a scalable amount." As an interactive graphic journalist, Journalist 6 said that the majority of the semi-automated charts he builds pertain to financial news, whether it be business earnings or market movements: "What I primarily do is financial stories, and it's particularly suited for automation because it's a lot of numbers. You know, quantitative data is a lot easier to deal with at the moment." Journalist 4, who had mentioned during the interview that this was still in a pilot phase, spoke about automating stock market news:

> The idea will be to do one story at the end of the day, we have someone that writes those now. They look at the NASDAQ, the S&P, the TSX, the dollar, gold, oil […] They do that at [market close] at the end of the day. Once automated, we can do that every hour. We can do that every half hour, or every 15 minutes.

Politics and Elections coverage

In terms of political stories and election coverage, most participants emphasized that they use automated tools to assist them in their reporting when writing these stories. Journalist 7 provided an example of a time he utilized data mining in order to "crunch a year's worth of sales from the federal government's surplus website to find out what the large ticket items were." By combining his reporting skills (i.e., finding sources) with pattern recognition software, he discovered that specific emergency vehicles funded by taxpayer dollars "were being sold for way lower than their blue book value." Using the same combination of data mining and reporting, Journalist 7 wrote another story about how the Ontario government had made millions of dollars in profit by breaking into an industry in the private sector. Journalist 4, who had built the election

news bot that operated through Slack, believes that this tool may not be ground-breaking, but that it adds more nuance and colour to his newsroom's stories than before:

> It's not a huge thing to know that the riding you live in is the fourth most densely populated in the entire country, but that's not something you would've known before, but it certainly adds context when you're talking in terms of housing, social policy, that sort of thing. So that's stuff that we never would have written before, but that our clients and readers now get […] It's not about gigantic increases, it's about marginal increases and I think yeah, it makes what we do better, provides more information, more context.

Perhaps one of the most unique examples of applying automated journalism to produce a news story came from Journalist 2, who utilized IBM Watson to provide an activist with an artificial intelligence-powered platform that allowed people to have a "virtual conversation" with the activist. "[The project] used machine learning and voice recognition to answer user questions with live edited video so that you could have a simulated conversation with her about her life, activism and country," he explained. In order to build a database of answers for the virtual activist, Journalist 2 spent two days interviewing her and asking a range of questions.

Crime Reporting

Another example of mixing traditional reporting with advanced automated technologies came from Journalist 9, who spent a year trying to find where crime guns in Canada were coming from. He explained that they filed over dozens of Freedom of Information (FOI's) requests to 36 different police forces. The project was a combination of developing and talking to sources, filing FOI's and using sophisticated analysis methods:

> What we learned was that it was an unanswerable question because the data just doesn't exist, but we did learn some very interesting stuff in the process. For instance, the number of entries in Canada's National Ballistics Database for shootings tied to assault rifles or assault-type rifles—which is something that the liberal government has said they're going to ban—is vanishingly small. It's so tiny that a ban on assault rifles, according to the ballistics database, probably won't have much of an impact on the number of shootings in Canada. So I think that's a pretty good example of using a combination of more traditional reporting tools like

> source building and filing FOI's with the more advanced stuff, like putting those FOI's into a data analysis pipeline and using some more advanced stats to decode what is actually going on inside the dataset. (Journalist 9)

4.3. Collective Self in Relation to Technology

The "Collective self in relation to technology" theme considers the ingroup-shared attributes of journalists (the collective self) in relation to automated technologies from the viewpoint of the interview participants. It outlines the participants' observations of the relationship between the collective self (media professionals and newsrooms) and automated technologies from a historical, sociocultural and economic perspective. In addition, it addresses how the collective self feels about the current state of journalism in Canada, and whether they view automated journalism as a requirement to remain relevant in today's media landscape.

4.3.1. Reception of technologies within media organizations

Journalism is a discipline where "communication has long been conceptualized as a human process through machines," where technology is the medium or channel that enables human-to-human communication (Guzman, 2018, p. 4). The intersection of technological change and news production often analyzes the ways in which technology, much like telegraphy and photography, have shifted organizational structures and journalistic practice within media organizations, while keeping journalists at the forefront (Anderson, 2012; Diakopoulos, 2019). This sub-theme gathers the participants' mixed feelings about how the integration of technology is received by journalists from a historical and socio-cultural perspective.

The majority of interview participants think it is challenging for traditional media in particular to accept incoming technological change:

> It's not easy. I would say that especially traditional media is having a bit of a reckoning now with the fact that [we] haven't really been nimble and able to incorporate tech easily because our processes are set in such ways. (Journalist 1)

When speaking of the tensions that exist between traditional media and technological change, Journalist 3 mentioned that he senses a resistance to the idea: "There are some that think that the old leather shoe reporting, where you just talk to people and look through documents should be enough to do the job." Others shared similar views:

> [Journalists] that have been [reporting] for years are resistant and I think are just wanting to hold on. Things have changed. They haven't changed that rapidly, but to these people in the industry, they seem to think [it has]. So when new things come along, if they don't understand them or it changes their workflow, or it could mean them doing something else—they're not super into that. You'd run into some resistance there. (Journalist 4)

> There is a skepticism now that technology in and of itself is not a good thing. I think things like the Dotcom Bubble, things like the temporary stature of things, like Snapchat and Napster—those kinds of fleeting technologies that were supposed to change everything and didn't—have left some people skeptical […] for example, TikTok is out now and it's sort of the hot new social medium and people are wondering, 'Oh, is this something that news needs to pay attention to?' And there's a lot of skepticism that it's just another one of these platforms that nothing's going to happen on. (Journalist 5)

In contrast, Journalist 8 was the only participant to argue that journalists are accepting of technologically driven change, as it is an industry in continual development, and new technologies constantly change the skills and jobs required of media professionals in order to produce news— a view shared by Deuze (2005) and Wilkinson & Winseck (2019):

> There's this myth that journalists are resistant to the adoption of new technologies, but they're not […] when we look at the history of journalism, there have been innovations and new tools that have appeared. There have always been early adopters of the telegraph and of photography even before that.

While Journalist 8 refers to the industry as a whole, Journalists 2, 4 and 5 are careful to attribute their observations to traditional media and older generations of journalists, as they—along with other interviewees—acknowledged that the "younger generations of journalists" (Journalist 7) are eager to use newer reporting methods. Journalist 4 believes that "[younger people] who just

got into journalism are really receptive to the idea of things changing. They don't think the internet is a fad. I think some of the older people do." Similarly, Journalist 7 doesn't consider the issue to be necessarily limited to journalism organizations:

> The people who grow up used to similar types of technology, for example the Gen Z's, would probably be more open to a new type of social media management platform than someone who's 60 and didn't grow up with the internet until much later in their career. So I think that younger generations are generally more receptive to technology in that they've been more immersed in it, or it's part of their normal lives.

In addition, interview participants suggested that journalists have a circumspect response to the adoption of new technologies—particularly within the scope of automated journalism—because "journalists are skeptical by nature" (Journalist 2). Three of the nine participants mentioned that this hesitation is particularly prevalent at the organizational level, given the general unbending nature of Canadian newsrooms:

> Obviously, [automation] comes at a cost, and it comes with training and workflow changes. So change management is very real in journalism right now and I think it's where a lot of companies are having a hard time. (Journalist 1)

> I think the Canadian journalism sort of mentality […] the culture is very risk averse. We don't like to take chances. We don't like to try things that haven't been tested and proven. We kind of wait for the Americans to figure it out, to take the risks and then it's like, 'Okay, that works. Let's try it.' And it's a long lag until it comes to that […] I don't know the inner workings of newsroom managers' minds […] but when it happens, there's no real strategy behind it and the expectations of what [automation] can do aren't very well defined. (Journalist 3)

Journalist 4 added that he really worries about stifling innovation: "I think that happens a lot, especially in journalism, and especially in Canada. [We're] not at a point where we can be slow about [innovation]."

By looking at the participants' comments above through Lewis & Westlund's (2015) sociotechnical lens, it is possible to recognize media's skepticism about the extent to which technological actants should be permitted to intersect with business growth, production, and

distribution at the management level. Here, the financial costs associated with these technologies are being discussed in relation to the ways in which automated journalism could disrupt how newsrooms are managed, and the subsequent fragmentation of certain work tasks. However, issues pertaining to change management when integrating these technologies might not be emphasized so heavily if the relationships between editorial workers, businesspeople and those specialized in building and developing these technologies were better connected in order to support the bottom line of cross-media newswork (Lewis & Westlund, 2015). This suggests that Canadian newsrooms might not yet be mediating "growing forms of cross-awareness and coordination between the editorial and business sides" (Lewis & Westlund, 2015). Journalists who can code, according to Journalist 3, "can be a liaison between the editorial and development desk." Furthermore, participant observations indicate that, while newsrooms increasingly carve out spaces for algorithms and intelligent systems, they are not keeping pace with its increasingly active role in finding, producing and distributing news. It is no longer sufficient merely to examine these forms of technology as passive mediums or channels for human-to-human communication (Guzman, 2018). Instead, media organizations must revaluate their business models while also positioning these technologies at the heart of their strategies, rather than containing them—and the various departments involved—in separate media silos, if they do choose to adopt it. This topic will be further examined in the next section of this chapter.

Moreover, these findings circle back to the discourse around journalists' ongoing fear of displacement, as intelligent machines and sophisticated algorithms become better able to analyze the exponentially available troves of data (Carlson, 2015; Graefe, 2016; Linden, 2017). Journalist 2 argues that journalists and newsrooms alike are skeptical, and justifiably so. According to

Journalist 2, new technologies should always be challenged when they risk impacting the norms and conventions of journalism:

> Some technologies have come with pretty hefty costs from the journalistic perspective. You know, on the one hand, it's great to be able to track your audiences, but if you're suddenly getting pressure from above to maximize based on audience counts, then you might have to sacrifice some very important journalistic concepts to get there. So, I think journalists are understandably wary of those kinds of trade-offs and I think we want them to be skeptical of new technologies and not just dive right in. (Journalist 2)

This view resonates with the shared journalism ideology among working professionals (Deuze, 2005), where journalists see themselves as maintaining a "strong moral and ethical foundation that gives journalists a legitimate and valid reason to act on behalf of the public" (Linden, 2017). As such, it would be logical for journalists to instinctively meet automated journalism with both resistance and mitigation. Journalist 8, who champions the shared professional ideology of journalists, strongly believes that "the core and foundational skills of journalism" will never need to change in order to accommodate automation:

> Being able to interact with sources, to go out and ask the difficult questions in the public interest at all times—those skills remain essential. So, even if a journalist doesn't necessarily know how to code or build robots, these fundamental journalism skills will always be quintessential to the industry's functioning.

During the interview, Journalist 2 also stressed that "core journalistic skills will never go out of fashion, but it's the fear of losing them [to automation] that keeps people from adapting." He added that "if you're an experienced journalist who's good at rooting up the truth and contextualizing information—those skills will never become obsolete. I just wish people were less afraid [of automation]." Journalist 1 reiterated this collective fear: "People are scared shitless […] because a lot of newsrooms have shrunk and have been chipped away at. They're scared AI is coming to replace them because that's where a lot of the wording around AI is." While these findings clearly indicate that "automation anxiety" (Akst, 2013) remains prevalent among media

professionals, several participants explained why they believe humans reporters will always be at the forefront of newswork:

> There's a lot of journalism that only a human can do and probably for the foreseeable future, will be able to do. I don't imagine a bot is going to write a really cool investigative story or a long feature that's going to tug at your heartstrings. I don't believe that's anytime soon. (Journalist 5)

> Machines will never have critical thinking because [machines] need to be taught. You can't teach something common sense. You can teach [a machine] to base itself on different factors, but that's algorithmic. A brain is programmed, sure. The way our neurons work […] it's not coded. So, I don't think—unless we completely do away with critical thinking—I don't think bots will be replacing journalists anytime soon. (Journalist 1)

> There's still a ton of reporting that has to happen […] because I have to figure out: what are the right questions to ask the data? Who can contextualize this information for me? Is this information even interesting or useful? […] How do I assemble that into a coherent narrative, because I can't just vomit a bunch of numbers on a page and hope that people will put it together in their minds. (Journalist 9)

Journalist 4 shared a specific example pertaining to his business bot, to assert his view that journalists are, and always will be, essential to the industry:

> [Say] the Canadian dollar is up 0.2 cents compared to yesterday. What my robot writes is that it's up or down. It doesn't say, 'Justin Trudeau's lack of fortitude with the Chinese government has hurt the Canadian dollar.' It's not making those types of decisions. It's really just giving facts. We haven't built anything—and I can't see us in the near future building anything—that has any sort of interpretation. I mean this goes back to what humans do great versus what computers do great, and computers do really great when it's just plain facts. Humans do really great when there's nuances.

Needless to say, the human vs. machine cognitive intelligence debate continues to be a dominant topic of discussion since it was first broached by Alan Turing a little over 70 years ago (see Section 1.1., Chapter 1). These findings suggest that participants share Dreyfus' (1993) view, wherein AI will never reach the mental capacity of humans, because intelligence cannot be reduced to mere symbolic manipulation (p. 20). The discourse continues to revolve around what are considered to be the core skills and values of journalism (i.e., creativity, critical thinking and ethical judgement). While all participants agree that automation is undoubtedly going to

eventually make its way into all newsrooms, less is said in terms of how this could lead to the upskilling or deskilling of journalists.

Nevertheless, one participant (Journalist 9) did say that automation has had a direct effect on journalists in his newsroom, suggesting that it was partially responsible for job loss:

> I mean, there are definitely fewer people working in journalism now than there were two, three, five, 10 years ago. I think automation is a big part of that. We have far fewer homepage editors [and graphics editors] than we used to. It's the same almost everywhere except for maybe the *New York Times*, which I think now accounts for something like 20% of all working journalists in the United States. Again, [newsrooms] are trying to do more with less, and automation—especially on the lower end of the types of tasks that people are performing, has freed up a lot of people's time, though it's also forced a lot of people into retirement. So, [automation] is not always positive. There's a lot of people that I knew when I first started out here who no longer work at [media organization] because of things like buyouts and early retirement packages—partly because their jobs are being automated or automation was having knock-on effects that were leading them to have to exit the industry.

Despite the job losses, Journalist 4 added that AI-enabled journalism is a benefit—even for those who don't practice it: "it's going to make our journalism better." He emphasized that it's not being integrated "to replace journalists, it's just to make the journalism that they do better." While he maintained that journalists, whether they know how to use these technologies or not, will produce better journalism thanks to automated journalism's capabilities, he also conceded that it has "forced people into retirement" and that its integration isn't necessarily a positive thing, thus eliciting contradictory opinions. Yet, some researchers, like Stray (2016), argue that "reporters augmented with intelligent software—cyborg reporters—have already pushed past the ancient limitations of a single human brain." This could mean that journalists without automation skills could become increasingly dependent on colleagues with these specialist skills (Linden, 2017), or in the worst-case scenario, out of a job. Earlier in this chapter, examples of increased collaborative efforts between programmers and journalists were provided by Journalists 4 and 5, although it is unclear whether this form of cooperation has been successful long-term, and

whether it has extended into other Canadian newsrooms. However, Journalist 9, who built his newsrooms' automatic chart generation tool, explained that the tool was precisely conceived with the idea that it would "democratize that kind of work in the newsroom so a reporter […] wouldn't have to go through three or four different people to get [a graphic] done."

While holding a more optimistic view, Linden (2017) believes that recent examples of collaborative data journalism projects demonstrate that the "sharing of knowledge, tricks and work practices" have now become some of the key values of computational journalism, and extend to the application of more sophisticated technologies. Even so, journalists without these specialist skills would require some basic understanding of these technologies if this type of collaboration was to succeed. Therefore, it is likely that the responsibility of training working journalists in this field will fall back on employers, which might be challenging for newsrooms. As mentioned earlier, automation comes with important training and workflow changes and "it's where a lot of companies are having a hard time" (Journalist 1). Yet, Journalist 6 doesn't think this is an issue, arguing that media organizations of today should already be much more data-driven, with regular training and upskilling integrated "into the culture of the company." However, Journalist 6's newsroom is well-known for being data-centric, which does not necessarily imply that other organizations have taken this step, since other participants expressed concerns for traditional media. Meanwhile, Journalist 5 remains hopeful that today's questions will provide solutions for tomorrow: "I think the fact that we're asking these questions now and that we're completely aware of the potential, I think helps us create systems that enhance jobs rather than replace or destroy them."

Furthermore, the axiom that automation equals better journalism remains somewhat paradoxical, as interviewees offered many diverging views about the effects of automation on

working journalists. Journalist 8 believes automation to be unavoidable, but holds a deterministic view, stating that "journalism won't be able to escape the automation of certain tasks and functions […] but it's more of an opportunity than a danger." On the other hand, Journalist 5 thinks automated journalism should not be considered a threat, but does understand the caution:

> I remember when robots were introduced into the auto industry. The idea was that they were going to help people [do their job], and then as robots got more advanced, they did actually start replacing people. And you know, there is a concern about that.

Two participants claimed that the introduction of newsroom automation could certainly become a threat depending on the types of stories or tasks a journalist conducts in his day-to-day work: "It could be a threat to people who only do very simple agenda-driven stories. Things that happen regularly and that don't take much analytical skill. You're basically just rewriting information that's put out there, rewriting out news releases," said Journalist 3. Correspondingly, Journalist 9 stated that "if the job is automatable, it was always going to be automated at some point." He continued with the following:

> When I say automatable, there's a lot that's packed into that, right? Like the question of whether the automated task is better than the human version of that same task, or the output of that task rather. So you know, if we're talking about things like generating charts, than the tool we built was basically the equivalent to what people were doing before that. (Journalist 9)

And even when talking about his own career, Journalist 9 seemed optimistic while also acknowledging that automation might eventually take over: "I'm pretty bullish about the future of automated journalism. I have no doubt that I will at some point be looking for a job, maybe partly because of automation […] but that's just the nature of the industry right now." As will be discussed in the next section, the participants' concerns are primarily related to the decisions being made at the upper echelons of the industry, and how certain business strategies could amplify automation's threat to human reporters.

4.3.2. New technologies and the journalistic business model

Some of the literature has maintained that automation is not the culprit behind industry decline (Smith & Anderson, 2014; Linden, 2017). Many of the issues that arise about artificial intelligence in journalism today, such as the threat to journalists' livelihood, are issues that journalists already face. In arguing that automation is not the driving factor, Linden (2017) writes: "Where journalism jobs have been disappearing, the reasons are to be found in changing consumer behaviour and media business models, not in automation, at least not so far." In the previous section, Journalist 5 stated that he was hopeful that the ongoing debate over the automatability of journalism jobs would help the industry lay out some groundwork to protect its media professionals. Despite this, he conceded that failing journalistic business models add to rising industry pressures:

> Newsrooms have shrunk. Not because of automation, but because the business model for news is falling apart. Time is short. People are doing a million things and if we can take some of that burden off of them, then they can actually do more valuable work that only a human can do. (Journalist 5)

Although the previous section presented diverging views among interview participants about whether automated newsroom technologies are a benefit or a threat to the journalism profession, the vast majority cautioned against the adoption of these technologies, especially if they are simply being incorporated to reduce labour costs and to replace staff. Journalist 3 said that it could become a threat, but that it could also relieve journalists so that they could pursue better work, "work that actually makes a difference." He added that the threat will depend on media organizations' objectives:

> I just hope that that's what a newsroom manager sees as well, instead of just, 'Oh, I can fire people and save money, right?' That's where it could be a threat, where there's a lack of imagination and cost-cutting is the only priority. (Journalist 3)

Indeed, the literature has extensively attested to the cost-saving capabilities of automated journalism as one of the underlying factors incentivizing its adoption (Carlson, 2015; Montal & Reich, 2016; Lokot & Diakopoulos, 2016; Graefe, 2016). According to Cohen (2015), declining profits and increasing demand for content have led media organizations to view automated journalism as a business opportunity to reduce production costs while increasing profit margins. This thesis presents the optimistic, complementarity frame found in the literature, and supported by discussions with interview participants, where automation is offered up as a way to free up journalists from routine tasks to allow them to spend time producing better quality journalism (van Dalen, 2012; Graefe, 2016). It also presents the more pessimistic outlook, where automation is only valued for its cost-saving advantage, which consequently threatens journalism jobs. Two participants in upper management positions made sure to stress that their newsrooms didn't adopt automated journalism to replace staff, but rather to keep up with rising demands:

> We haven't really had a choice to maintain our output with reduced staffing. Now I want to make it clear though, that in the entire conversation we've had at [media organization], the goal of AI was never to replace staff. That's been a big ethical standpoint that we've talked about. It's never to replace staff, it's just to help them do their job in a different way. We need it to fill some holes in different places or just to assist them in the tasks that they're doing so that the output doesn't get affected. (Journalist 1)

> I can't speak for every news outlet, including the ones that are already using automation, but it's never been my goal or my [media organization's] goal to bring automation in so that we can reduce jobs. It's not a cost-saving measure. What it does save is journalists' time […] So, if you work for a reputable outlet, that has journalism at its core, I would say that there shouldn't be anything to worry about with automation. I suppose if a company is employing journalists only to do that stuff—turning a box score into a hockey story—then automation can replace them. (Journalist 5)

According to Journalist 4, who also holds a management position, replacing journalists with automation will ultimately hurt newsrooms in the long run, because human input and editorial oversight remain essential:

> I don't see any major disadvantages, unless companies decide that they're going to try to replace reporters with this. I know that's not what we're doing at [media organization] and I've made that very clear. If other places tried to do that, you know, good luck to them. There still needs to be a human element involved in these sorts of things. And if [they] just try automating everything, it's not going to work out for them.

Journalist 7 reiterated a similar view, cautioning against this sort of strategy:

> [Some newsrooms] think that this might be a way to save costs by getting rid of journalists. [In my opinion] that's a falsehood, but I worry about that perception […] because what that means is that there might be people who want to adopt this technology to downsize journalists erroneously.

Although interview participants outlined the various ways algorithms have helped advance news production, they also acknowledged the tensions journalists increasingly face to publish viral stories: "we need more content, we need more stories. That's where the business model is now. We need more clicks, so we have to produce more stuff. We've also been victims of cuts, we've seen that," said Journalist 1. When Journalist 4 built an automated tool that lets reporters add a watermark to their tweets, he mentioned that his newsroom saw more "engagement on their [Twitter account] that way." As aforementioned, Journalist 2 relayed this notion to how this makes normative standards and ethical work practices harder to follow: "if you're suddenly getting pressure from above to maximize based on audience counts, then you might have to sacrifice some very important journalistic concepts to get there." This notion is reinforced by Noble (2018) in her book, *Algorithms of Oppression: How search engines reinforce racism.* In it, she states that journalists "are facing screens that deliver real-time analytics about the virality of their stories. Under these circumstances, journalists are encouraged to modify headlines and keywords within a news story to promote greater traction and sharing among readers" (p. 154). Journalist 1 expanded on these rising tensions as algorithms encroach his newsroom and other media outlets:

> Now everyone will publish [the stories] that drive the most engagement, which drives clicks, which drives money. And so that kind of sucks, right? So algorithmically, I think there's a huge issue with the way we decide to cover things […] I think a lot of places are letting go of covering certain topics and certain spheres because it wasn't driving engagement.

Yet, a recent shift towards full-fledged automation by Microsoft typifies exactly what the interviewees have been cautioning against. In May of this year, Microsoft announced that it was to replace about 50 of its part-time news production contractors with artificial intelligence-enhanced systems. An article published by the *Seattle Times* reported the following statement from Microsoft: "Like all companies, we evaluate our business on a regular basis. This can result in increased investment in some places and, from time to time, re-deployment in others" (Baker, 2020). The work that was previously done by the employees was described as using algorithms to select and curate trending news stories from various news organizations, and content optimization. As one terminated employee told Baker (2020), the job has been "semi-automated for a few months, but now it's full speed ahead. It's demoralizing to think machines can replace us, but there you go." The work responsibilities held by the former Microsoft employees—aggregation, content optimization, and curation—fall in line with the types of routine tasks that the literature and some of the interview participants have deemed to be automatable, but it remains to be seen whether the machines will do a better job than humans. "Some of these menial tasks are still at the core of what can make you a great journalist and where you get to learn and kind of cut your teeth," stated Journalist 1. He continued by referring back to his newsroom's translation desk, as discussed in section 4.2.3. of this chapter:

> Translating one hundred stories from scratch takes a lot of your brain and mental capacity. If all of our translators were to just do 100% editing because the machine was just spitting out good content, they wouldn't be translators anymore. It's not a huge disadvantage […] but I think we have to look at what is worth automating and what's not […] Because if you're just automating for the sake of automating without really knowing how it's going to impact your bottom line—then why are you doing it? But if you know that it's core to the work that you're giving to people or to your business model, then there is a benefit.

What happened at Microsoft drives home a potentially critical flaw in the paradigm of the human-machine hybridization championed by the likes of Diakopoulos (2019) and Woolley, Boyd & Broussard (2016), namely that journalists are safe from job cuts while algorithms and intelligent machines take care of their more routine tasks. However, a panacea exists in the incorporation of updated communication and collaboration strategies between the different actors at play that would "reinforce the complementary nature of the technology" (WAN-IFRA, 2019),[6] which in turn could bolster the journalistic values that make up the collective professional identity (Lewis & Westlund, 2015). In WAN-IFRA's 2019 report on automated journalism, the executive summary reads as follows: "the good news is that so far, news automation has not replaced humans, and looks set to work alongside humans in the newsroom."

Yet, the year the report was published, Journalist 9 partially attributed job losses in his newsroom to automation. Less than a year later, Microsoft fully replaced dozens of its workers with artificial intelligence-enhanced systems. However, a recent error made by Microsoft's automated programs will be discussed in detail later in Chapter 5. These examples attest to the rapid pace at which automation is evolving and being incorporated into newsrooms in Canada and worldwide. As such, if automated journalism is to be increasingly integrated into Canadian newsrooms, it becomes imperative for business and editorial strategies to be put in place in order to maintain the routines, skills and values that make up the professional identity. The talks with the nine media professionals point to what seems to be a patchwork of solutions set in place in newsrooms currently exercising automated journalism across the country. Although all interview participants expressed that automated journalism can be complementary to human journalists,

[6] WAN-IFRA is short for World Association of Newspapers and News Publishers, a news consortium that published a report on automated journalism in 2019.

they worry that at the organizational level, automation is little understood, which could lead to serious competition and a threat to the security of journalism jobs (van Dalen, 2012). If this area of journalism continues to grow, so should the holistic, editorial and business strategies.

4.3.3. Motives and limitations

This sub-theme explores the motives behind the adoption, or rejection, of automated journalism by Canadian news outlets as perceived by the interview participants. It also examines the findings pertaining to the state of automated journalism in Canada within the context of those motivating forces, and what participants consider to be the current limitations of these technological tools and processes, including limitations in its uptake.

As aforementioned, Canadian newsrooms continue to be confronted with ever-growing pressures to extract higher profit margins with the little resources they have left at their disposal due to failing business models. According to the literature, these are the principal factors driving automation, as the goals of newsroom automation are to free up journalists from spending time on menial tasks, while increasing output and expanding coverage at lower costs (Broussard, 2015; Carlson, 2015; Cohen, 2015; Montal & Reich, 2016; Lokot & Diakopoulos, 2016; Graefe, 2016; Stray, 2016).

All nine interview participants mentioned that their newsrooms, or they themselves, implement automated technologies because it enables them to keep up with increasing demands for journalistic output, to ease workflow, and to expand coverage. Several participants noted that another motivation behind its adoption is simply newsroom innovation. "[Our motivation] is just the sheer idea of innovation and trying new stuff—thinking outside the box. I think it's foolish not to try new technologies [for fear] that they might fail," said Journalist 1, who also added that

his newsroom mandate is innovation and bringing in new tools. Similarly, Journalist 5 said that "the possibilities [with automation] are really limitless right now. We're just scratching the surface." For Journalist 8, automation opens up new doors for the younger generations of journalists:

> Technology has always eliminated jobs, but it has created two, three, ten times as many. It's an opportunity to introduce new skills to up-and-coming journalists, that will allow them to push the boundaries of journalism even further. It leaves room for innovation, and creates jobs that probably don't even exist yet.

While "innovation" is discussed above in a broader sense, Journalists 4 and 5 had the most interesting examples, as their newsrooms are exploring how they could obtain new sources of revenue with the help of automated journalism. "One of the big things I'm really pushing is that we need to find new avenues for revenue," said Journalist 4, while admitting that this is mainly driven by the dire state of journalism in Canada. He explained that when he automates 10 hours out of someone's job, he expects them to "spend those 10 hours trying to come up with a new product or service that we can sell, instead of just producing more content for the wire." Although he declined to go into detail about the financial opportunities currently being explored, he did emphasize that he felt a sense of urgency given journalism's financial situation:

> The industry is in a pretty dire sort of state right now. We need to try a lot of new things really quickly and see if we can get anything to work. I'm not saying that AI is the silver bullet to do these sorts of things, but it might be a way [for us] to tread water for a while [or] develop different ways of generating revenue. That's where my thinking is. Let's automate some work and then use the free time that we have to find a new way to make money.

On the other hand, Journalist 5 mentioned that his newswire service is looking at potentially expanding its client base to companies outside of journalism:

> I mean, we've talked about other sports possibilities and other opportunities to sell our content to outlets other than news outlets. There may be professional brands, businesses or other parts of industry—like academia or government—who would want this kind of information that we're now doing with automation, as opposed to just [selling to] newspapers, TV, radio, online websites.

Given the discussions with all the participants throughout the interview process, it comes as no surprise that certain newsrooms, much like the ones employing Journalists 4 and 5, would turn their efforts toward improving their financial margins with the help of their automated tools and processes in place. Based on the interviews with both journalists with regards to the types of automated tools being implemented, and the tasks they help accomplish, their newsrooms appear to be some of the most advanced in the country, and certainly compared to the ones employing the other interviewees (with the exception of the freelancer).

As previously mentioned, the bulk of the research highlights the low marginal costs as one of the main benefits of automated journalism (Graefe, 2016; Diakopoulos, 2019). However, this advantage of automation only accounts for when it is fully integrated and operating within the news production pipeline. Limitations to the adoption of automated journalism tools and processes include important costs. This was purported by several participants throughout the interviews, from the cost of hiring specialists with the necessary skills, to the expenses involved in building and upkeeping these systems (WAN-IFRA, 2019). Journalist 5 indicated that for his media organization, adopting automated journalism was not a "cost-saving measure."

Natural language generation (NLG) systems, much like the one in place for the junior hockey games recaps in Journalist 4's newsroom, require a lot of upkeep and according to WAN-IFRA (2019), NLG systems "require a lot of expensive development work." Journalist 4 added that there are only a few companies in Canada that have adopted automated journalism, not only because "it is certainly not the norm yet," but because "it's generally really expensive." In order to adopt these types of technologies, Journalist 6 put it bluntly: "well, you need money," and while discussing the state of Quebec journalism more specifically, he added that it's in "a bit of a bad financial situation […] which always has to do with money and resources," hence why

automation hasn't really been adopted. Journalist 1 spoke of the important expenses associated to hiring dedicated specialists:

> Hiring dedicated programmers to build and upkeep these systems comes at a crazy cost. Unless your [newsroom] is willing to go ahead and fork out that money, outsourcing or external funding becomes kind of necessary […] there's no way in hell we were going to get our automation projects done [without external funding].[7]

Hiring dedicated experts proves to be a challenge for several reasons in Canada. For one, automated journalism can only become fully integrated if media organizations can allot the necessary funds, either in hiring a dedicated team to build these tools and processes in-house, or in negotiating with external actors. All nine participants have pointed out that the Canadian industry is in a dire state, and attested to the fact that their newsrooms lack resources, making it unlikely that this will change in the near future. "It requires engineers and we don't have the money for that […] So until we see drastic investments by Canadian media companies in engineers and tech people, I don't think Canada will be driving the industry," estimated Journalist 2. Journalist 3 echoed similar sentiments by adding that Canadian newsrooms aren't hiring specialists, "due to a lack of vision, lack of resources, and a lack of people in leadership positions who understand the value of this."

The main factor limiting automated journalism's widespread adoption then, falls back on what participants have said about the industry's failing business model. Without readily available funds, newsrooms can't really push automation forward, and although external funding opportunities exist, they remain limited. Journalist 6 mentioned that one of his former jobs, where he filled a role similar to his current position, was subsidized by Google: "If it wasn't [subsidized], I wouldn't have had a job, I wouldn't have been doing journalism."

[7] Journalist 1 is referring to the $300,000 in external funding that was awarded to his newsroom by a large tech company so that they could get their automation projects off the ground.

Secondly, there are very few journalists in Canada who are specialized in this field. Not only was this apparent while this thesis was in its initial recruitment phase, but participants also expressed that this was part of their own working realities. Journalist 3, stated that: "we are a very, very tiny group in Canada. We used to have a Slack channel, but it died down because nobody was talking […] we didn't have the critical mass to get going." Journalist 4 said something similar: "There's very few people in Canada […] I'm sure you haven't really found a bunch of people in Canada for your thesis, because no one is really doing this. The Canadian industry is not there." And while Journalist 7 shared the same views as the other participants, he was the only journalist to address the increasing demands made by newsrooms for qualified people to fill these positions:

> I mean, you could be really experienced, but are your skills compatible with today's news market? And we're not there yet, where I think everyone has to have those skills, but I think we're getting there […] being able to interpret large amounts of data is kind of necessary as more and more of the world is generating data. You have to be able to report on the world the same way you used to, but with speed and scale […] and with less people.

Yet, rising demands for specialist individuals run counter to the troubled industry being described. As will be discussed in the next section, the very nature of automated journalism demands a new form of editorial oversight, and a lack thereof is problematic, seeing that newsroom algorithms must be held accountable. Not only because the inherent biases built into them "can skew public perception," but also because media organizations are aiming to build their newsrooms' identities (i.e., deliberate values and intentions) into these systems (Diakopoulos, 2019, p. 234 & p. 241). "There are very few people who work in newsrooms, in management who are qualified to vet these things […] There's no one who can do this and that's a problem, and something we need to figure out," cautioned Journalist 3. However, Journalist 5

believes the solution to already be underway thanks to journalism schools who are making space for this type of training in their curricula:

> Quite a few [journalism schools] are getting people into code—understanding how to code or what algorithms mean. So I think there might be more journalists coming down the pipe that understand what's under the hood of these automated systems.

Journalist 8 also mentioned that his second-year students are required to take a class to learn how to code: "Learning how to code isn't just nice to have. To me, it should be part of a journalists' toolbox."

4.4. Technology as the Professional Self

The final theme, "Technology as the professional self," discusses the socio-cultural implications related to the emergence of technology as the communicator, whether journalistic content is fully- or semi-automated. It explores a range of ethical issues surrounding this technology as addressed by the participants, including: transparency, testing and verification, newsroom policies, and copyright and privacy issues. Ethics in journalism, and more broadly, within the field of communication, has long been grounded in the assumption that humans are the sole communicators (Guzman, 2018). The significance of automated journalism and its diverse functionalities must be considered, as it ultimately disrupts preconceived notions about the journalistic code of ethics, and subsequent work practices related to the use of this technology (Lewis & Westlund, 2015; Guzman, 2018).

4.4.1. Ethical implications

This sub-theme presents what policies have been put in place by newsrooms and how they vary. It also addresses policy gaps and ethical considerations which have not been

standardized across Canadian media organizations and where interview participants see room for improvement. Furthermore, it explores how, if at all, Canadian newsrooms are shifting their strategies to ensure that transparency and accountability measures are being put in place to accommodate these new forms of newswork.

Transparency

A study by Montal & Reich (2017) explored the perceptions of algorithmic authorship and the attribution policies set in place within newsrooms. They found major discrepancies concerning bylines and full disclosure policies regarding automated content. Similarly, this thesis' findings indicate that byline and disclosure policies are not well-defined in the newsrooms employing the interview participants. In discussing his newsrooms' byline policy surrounding the automated hockey games recaps, Journalist 4 said that "nowhere in these stories do we say that this was automatically generated by our bot, but the name of the person who vetted the story doesn't go on it either. It just says [media organization], but I don't get to make those types of decisions, unfortunately." In addition, Journalist 4 was the only interviewee who didn't think adding a byline was necessary:

> I'm not entirely sure to tell you the truth. The robot was programmed by a journalist. I don't think you necessarily need to say, 'Oh, this was written by a robot.' I'm not sure what the difference is. It's still something that was the product of someone with journalism training, or maybe not journalism training, but someone who's been vetted by the journalistic organization to put that robot out there. To me—maybe it's different because I work for a wire service—bylines don't bother me that much.

Meanwhile, Journalist 2 attested to the fact that when it comes to algorithmic authorship, "different [newsrooms] are doing things differently. I'm an advocate for transparency on that, but I don't know what [others] have decided to do." Journalist 5 also emphasized the importance of this form of transparency, whether a story is automatically generated or written by a human:

> [Automation] hasn't changed our transparency guidelines. We're very clear about [how] our content is produced. For our automated texts, there's a tag at the bottom […] that lets people know that this is powered by automation. So as we go forward with our digital data desk, all the content produced will have that kind of tagline. It's an essential part of journalism, where you have to let [your audience] know who wrote this story. And if the 'Who' becomes a 'What,' a robot, I think people deserve to know that. So, we're going to maintain that transparency.

Journalist 8 reinforced the responsibility that media organizations have toward their audience, a notion which was also discussed in Montal & Reich's (2017) study. During the interview, Journalist 8 said the following:

> The public needs to know who is talking to them. It's the basic principle of journalism. That's what I tell my students who are scared to go into broadcasting. They say, 'Oh! I don't want to see my face on TV' and I tell them it's not their faces that matter, it's the content [that's broadcasted]. If automated processes are involved in the generation of an article, the public must be made aware.

Journalist 1 said that his newsroom is "always transparent in terms of how a story was created," whether his newsroom publishes a story that's strictly AI, or reviewed and edited by a human. He added that "it's super important to be clear about the fact that an editor has gone over a story."

This leads to another aspect of journalistic transparency, which involves accompanying a news story with the methodology that outlines the automated processes involved and how data was retrieved and analyzed. Research on algorithmic authorship has revealed that alongside the issue of bylines, there is a critical need for algorithmic transparency concerning the "methodology, construction and limitations of the algorithm" (Montal & Reich, 2017). Broussard (2016) argues that this should be a standardized practice for any newsroom exercising forms of automated journalism, but this thesis' findings reveal that this is not the case at the media organizations employing the journalists who participated in this study. Two interviewees

admitted to inconsistently disclosing their methodologies. For one, Journalist 7 said that he only recently started incorporating this level of transparency in his reporting:

> I think you have to. There have been stories where I haven't really published my methodology. I don't want to be called a hypocrite, but at least recently I started publishing methodologies or plainly stating how I did something. I definitely think that we should all be held, and myself included, to that higher account of posting methodologies.

Journalist 9 believes that methodologies should only be incorporated for certain stories, even if all of them involved "advanced statistics and machine learning," because of its time-consuming nature:

> I mean, there's a bit of a threshold beyond which a methodology is necessary. I don't want to write a methodology piece for every story because it would be a huge pain in the ass for me, but also because it's extremely boring to almost every reader.

On the other hand, Journalist 8 refuted the work practice exemplified by Journalist 9 and argued that divulging a methodology is a requisite: "I think for an audience to have confidence in what we present to them, we have to be transparent about our approach. You have to be transparent at all times, because audience trust comes from transparency."

Indeed, some experts have argued that media organizations should inform their readers about how their material has been produced, arguing that they could "benefit from this form of transparency and explainability when competing for audiences" (WAN-IFRA, 2019). These deferring opinions reflect a larger issue at hand, which is the lack of policy concerning when and how to disclose where a story's data comes from, or how the analysis was conducted. "Methodologies are something that I do, but [my newsroom] has no policy that says that I have to do it," said Journalist 3. If some form of automation was used to assist in the actual reporting, "then obviously you need a methodology section that lays out everything you did in annoying detail," asserted Journalist 2. Journalist 2 ascribed data analysis to being one of the biggest changes to journalism "with the advent of so-called data-journalism." He explained that "instead

of having scientists working out of universities producing these analyses that journalists then write about, we're in-housing the analysis. We have to be transparent about how we're doing that." He added that before, "it was transparent because [the analysis] was in a research paper done by a scientist or academic somewhere and people could trace that back, but if we're doing it in-house, well, we better make it transparent."

As aforementioned, Journalist 9 does not always include methodologies in his reporting. "It's extremely boring to almost every reader," he said. Similarly, Journalist 3 stated that, although he always includes a methodology, "very few people actually read it. It's a very small proportion of people, but it's still good practice." He added that:

> If journalists are going to demand transparency from people in power, I think it's only fair that we be transparent ourselves with our processes. If we use data that we gathered and created, people should be able to use it or challenge us on the methodology. But like I said, it really hasn't bubbled up to the level where this is actual policy yet.

Nevertheless, while not many people read the methodologies that accompany a story, WAN-IFRA (2019) indicate in their report that "this does not necessarily mean a carte blanche" for media organizations. Still, Journalist 9 mentioned that his criteria for deciding whether a methodology section should be added or not falls back on how it will be received by an audience:

> I try to keep [methodologies] for the stories that take a long time, that are much more involved, both for me and for my editors, where it's going to get the kind of prominence that means that readers are going to be asking the question of, 'Okay, but how did they do any of this?'

Meanwhile, Journalist 4 was the only participant who stated that no methodology whatsoever accompanies his newsrooms' automated business and hockey stories. For the hockey stories, he explained, it's just "this team had more goals than that team. I don't think that sort of thing is up for debate. There's no interpretation there […] it's just plain facts." Although he added that if his

bot was writing more complex stories, "[adding methodologies] would be a topic up for discussion. We've never had to deal with that because we don't build [bots] that can do those types of things." Still, the hockey scores and the numbers that go into the sports or business stories come from an external source, thus indicating a lag in disclosure in his newsroom's practice involving automated processes.

It is important to highlight that Journalist 9 believes his newsroom to be pretty "ivory tower-ish" when it comes to their reporting. He conceded that, "we're going to need to start having conversations about publishing entire datasets, publishing code books. [We need to try] to be a bit more open about the steps that lead up to a story in the first place." Journalist 6 echoed a similar perspective about his newsroom, saying that "[they] are very protective of its data." It is plausible then, that newsrooms are reluctant about this form of full-fledged transparency for two reasons. One, there is certainly no established rulebook in Canada regarding methodological transparency for journalism and two, because of the competitive nature of the industry.

Copyright and data privacy concerns

Another important object of discussion was concerned with where data gets extracted from. This led to more discussions with interview participants about data transparency issues. More specifically, what websites a newsroom legally has the right to perform web scraping or data mining on, and to what extent media organizations should make this data available to an audience.

> [For Canadian journalism], we haven't really established good data ethics around the process of downloading or taking data. For example, if websites say 'You can't scrape us or we'll sue you.' Do we abide by those rules or not? Are they legally binding? Is it ethical to do that? I personally operate on the opinion that if something is a government database or source, then it's open season because it's paid for by tax dollars, but if it's a private listing,

> say you scrape Airbnb's website or a paywall and you're mining it, that's essentially copyright infringement or stealing. (Journalist 7)

As for Journalist 6, he said that his newsroom consults with a legal team on a regular basis when accessing data for a story, although this was done after the fact in the example that follows. In February, he was covering the number of Coronavirus cases in China and scraping a Chinese Web source that aggregates the number of cases per Chinese province. He further explained that ethical issues were raised about this process:

> The health departments of every [Chinese province] would announce their new number of cases daily. It was really long and hard to have to go to every province, to their health department website, extract the numbers and make them comparable. So what we did was take the feed from the aggregation website, collect the numbers all at once and put them into our graphic with a link under the graphic towards where we got the data. We didn't know we couldn't do that. We had a lawyer check it and he told us that it was illegal. Since our readers pay for our product, we aren't allowed to take this data, which is then "technically" being sold through us.

Although the growing availability of data is prompting more media organizations to implement automated technologies in order to extract, mine, analyze, and subsequently publish data-centric stories, data privacy issues remain unclear in journalism, according to Journalist 8. He mentioned that he was building his own research database using SEDAR, which is the electronic system for the official filing of documents by public companies and investment funds in Canada:

> SEDAR reached out to me and said that the data belongs to them, even if it's considered public data. They stated that if I wanted to automatically scrape their site, I had to pay $10,000 annually. I had to abandon my project because I was told that I was going to get sued. So you see, there are many problems with automation and here, I was about to be charged thousands of dollars even though I was building this project to do some reporting in the public interest.

These examples raise important considerations for newsrooms who wish to increasingly use this type of data, whether the data is considered to be public or proprietary information. Media organizations need to have people at the management level that understand how these different websites legally operate, and choose to move forward accordingly, while maintaining

their journalistic values. Although data may be publicly available, as is the case with SEDAR, perhaps this does not necessarily imply that it can be legally downloaded in full and distributed (WAN-INFRA, 2019). Journalist 9 provided another example that held ethical implications:

> A couple years ago, I helped a few colleagues on a story about some irregularities in voting within the Progressive Conservative Party of Ontario, for who was going to be the candidate or the nominee for certain ridings. That story required me to look at actual registered voter rolls for a few ridings. And of course, we can't publish that information. It's all private. You know, we didn't get that through a wire or anything. So there are also considerations like that, like, where do you draw the line? We can't. It's the same reason why we don't just publish document dumps like Wikileaks used to do because there's probably personal information in there. We wouldn't be responsible if we just put it out in the open. So there's also those kinds of considerations in all this stuff. It's a very murky field sometimes and you kind of have to just go with your gut.

Fact-checking and testing

The majority of participants strongly emphasized how important it is for newsrooms to have rigorous testing and fact-checking systems in place. Journalist 4 said that "we do a ton of testing before we release any [product]. We try to find every single edge case that we can to make sure our product works and it takes a lot of time." Journalist 6 explained that they do "unit tests, integration tests, regression tests, all the testing you can do, we do." Others shared similar examples:

> A few years ago, we started actively having a multiple fact checking system for code. We don't do that for every story because it would take an absurd amount of time to do that. But for larger stories, absolutely we do that. Whenever I work on a big story where the analysis is even a little bit complicated, I will get someone else to take a look within the newsroom who knows the code and knows enough about analysis to say like, 'Oh, why did you make that decision that's a bit weird.' We definitely have a code fact checking process now, but I don't know that a lot of other newsrooms do that, especially in Canada. (Journalist 9)

> We're very rigorous about testing our algorithms and making sure they're doing their jobs right. So, we spent a lot of time with our bots doing that. I think we did three months of testing to make sure it was rock solid and that also included working with journalists to make sure the wording of our [automatic text generation templates] was proper as well, because

> we wanted it to read, you know, fairly human. It should be hard to tell that a robot wrote this. So those are two kinds of safeguards we built. (Journalist 5)

Still, fact-checking data can become difficult if a newsroom lacks the qualified experts (see section 4.3.3.). For Journalist 3, "algorithmic accountability is a big issue these days," adding that, "I know very few journalists who can do that, people who go through the algorithm to try to reverse engineer it—input goes in, output goes out—but what's happening in the middle? How accountable are newsrooms being?" Journalist 1 expressed that this is where policies need to be considered:

> I think that's the part where I'm the iffiest about using AI. It's because it's created by humans. It's not fail-proof. We need to adopt fact-checking strategies and ethics. Those ethical guidelines need to be written down and everyone needs to be aware of them.

While hiring dedicated technologists has been presented throughout this chapter as being a challenge for Canadian newsrooms for various reasons (i.e., lack of resources, lack of qualified individuals), this also proves to be a hindrance for media organizations who have adopted automated journalism, but lack the ability to test and fact-check the veracity of their data and the automated systems running the analyses. According to Journalist 7:

> There's a lot less, at least at this stage, opportunities for editorial oversight because there aren't a lot of people in a newsroom that can write or interpret a machine learning algorithm, necessarily. Not that they're not capable, but they just don't inherently have that skill. It's not like copy editing or fact-checking in a standard way. And I think that being able to interpret large amounts of data is kind of necessary as more and more of the world is generating data.

Despite this, all interview participants underscored the fact that to their knowledge, everything that they or others in their newsroom produce using automated methods goes through some form of human oversight. "We never let something go that's strictly AI to our feed. There's always an editor that goes through it to verify the data, just to make sure," said Journalist 1. Similarly, Journalist 4 stated the following:

> There's always human intervention in everything that we do that's automated. So it's still the responsibility of the reporter, editor, whoever it is that's putting the information out there to ensure that it's accurate and that it's right, that there's nothing libelous.

Accountability

Overall, the participants specified that algorithmic accountability must always fall back on the journalist, or at the very least, the newsroom putting out the story. According to Journalist 5, technology shouldn't change the principles of journalism. "I think that the kinds of principles that trickle through journalism will transcend technology," he specified. Although he did not specify how his newsrooms' bots and algorithms—or the ones developing and maintaining them—are held accountable when mistakes occur. Perhaps this is indicative of the nascent state of automation in his newsroom and elsewhere, as others indicated the urgent need to revaluate journalistic policies to ensure algorithmic accountability:

> There's a whole new chapter that needs to be written and added to the Canadian journalism code of ethics. It should state that a human must always be involved, because if an automated process makes an error, but the story gets published anyway, who is liable? Who is responsible for that error? (Journalist 8)

> Well, you can't hold an algorithm accountable. You have to hold the journalists accountable and that's why, you know, your question earlier about authorship, it becomes even more important. I think that we have to not only credit the journalist, but credit the people who are programming these algorithms because, if there is bias in there, they need somebody to be answerable for it. (Journalist 2)

Interestingly, two journalists (1 and 4) shared very contrasting views about the adoption of newsroom policies to hone in automated journalism. Both expressed concern about the state of automated journalism in Canada, although for very different reasons:

> My concern right now is that—and that goes for the entire world, it goes for government, it goes for businesses—a lot of places are really quick to adopt tech and the policies come after. I think it should be the opposite. I think we should protect ourselves with policy and kind of a fail-proof thing, where [accountability strategies] are set in place before its adoption. I'm going to do a parallel with vaping. We're seeing all these heath issues and now

> it's a health crisis and we're like 'Oh my God, please stop vaping.' Why didn't we have pre-established rules about vaping, or at least like health guidelines about vaping? We're 10 years into vaping and only now realizing that it's bad for you when like, we knew that already.

On the other hand, Journalist 4 worries that if Canada focuses on developing policies to frame the integration and application of automated journalism before actually trying it out, it will stifle innovation, something he expressed concern over throughout the interview:

> They should eventually adopt policies, but I wouldn't put the cart before the horse. I don't think we're at the point where policies should be put in place yet. Once it becomes a bit more used I absolutely think it's something we should talk about, but in the meantime, I really worry about stifling innovation. Like if you have an idea, like adopting automated journalism, but your newsroom says 'No, no, no, we have to come up with a policy.' It stops the innovation from happening and we're not at a point where we can be slow about these sorts of things.

Moreover, while the nine interviews shed light on the various ways automated journalism is being implemented in their newsrooms, less was said about the policies and strategies that should be incorporated to reign in these technologies. This suggests that within their respective newsrooms, the discussions aren't so much about the implications of automated journalism for the future of newswork. Instead, they revolve around their failing business models and the urgent need to put financial strategies forward. "Journalists are a funny bunch because we're extremely allergic to uncertainty," said Journalist 3. And while the future of journalism seems uncertain, little is being done to thoughtfully consider the place of automated journalism in the increasingly competitive Canadian media landscape.

4.4.2. Audience engagement

This sub-theme evaluates the participants' perceptions of audience engagement, and whether automated journalism shifts the relationship between the reporter and the public as these technologies increasingly take on the role of communicator. It provides an overview of the nine

participants observations about the ways in which automated journalism might alter the way news is interpreted by a general audience, and whether there is an added value when increasingly adopting these tools. To a certain extent, this section attempts to fit audiences within the interplay of journalistic actors and technological actants, as the three entities are mutually engaged from a business and journalistic perspective in driving the way news is produced and subsequently consumed (Lewis & Westlund, 2015).

There was consensus among participants that automated journalism increases the span and breadth of their journalistic work, which in turn allows them to reach new audiences. This was exemplified by Journalists 4, 5, and 7 who spoke of the ability to expand local coverage thanks to automation. While Journalist 2 argued that "human beings are better at communicating with human beings," he acknowledged that for local news, "we just don't have the manpower and I think [automation] has been somewhat of a stop gap for the devastation of local journalism." He added that:

> When all of these local journalistic organizations are financially dying, that means a whole bunch of local town halls and local courts are not being covered and you know, a national news organization is not going to be able to make that up with manpower […] maybe you can create stop gaps using automated writing, but not to replace local journalists. For now, it's better than nothing.

Much of the existing automated journalism research attested to the fact that machine-written news has proven to be competitive with that of human journalists (Graefe et al. 2018; Wölker & Powell, 2018; Lewis, Guzman & Schmidt, 2019). However, none of the participants who use more advanced forms of automation, like automatic text generation, elaborated on whether their newsroom have explored analytics pertaining to user engagement with their automatically generated texts, and whether they have added value in driving audiences to their websites. Given that the majority of participants stressed that their newsrooms were feeling the rising pressures to

produce more content with fewer resources, it was surprising to find that this was not being evaluated from a business perspective. Still, Journalist 9 speculated that the actual "analytics and statistics on how often [our automated] articles get read are probably pretty dismal," this discussion also included how often methodologies get read, as mentioned in the previous section.

In terms of using AI and newsroom automated processes to directly engage with audiences, Journalist 2 referred back to the artificial intelligence-powered platform he created to allow people to have a "virtual conversation" with a transgender activist: "it creates new experiences that didn't exist before. It simulates a conversation with a real person. That is a new format that is made possible by machine learning." Journalist 1 added that:

> Using AI to communicate, in terms of using a robot messenger, is beneficial. If audience engagement is automated in parts, I think that's a huge benefit because a lot of people don't read news anymore and are seeking new ways to consume information, or to be informed on their own terms.

In contrast, Journalists 8 and 9 saw communication bots (ChatBots) as not necessarily benefitting newsrooms. For one, Journalist 8 said that he doesn't like the idea of using ChatBots to interact with the public, adding that it raises "many ethical questions." In sharing his personal opinion on the subject, he said that:

> I don't know many people that like it when it's a robot that speaks to them. If media organizations start automating their interaction with the public, I don't think their audience will be very receptive to that, at least not for now. Maybe these things will evolve and it will eventually happen, but again, they'll need to be transparent about the fact that it's an automated software engaging in communication with the public. For now, I don't sense that there's any real appetite for that, but I could be wrong.

On the other hand, Journalist 9's newsroom experimented with a social media bot that would interact with users and suggest stories around a given topic. "I don't know what happened with that, but I think it was a short-lived kind of project that we played around with. The appetite for

that stuff in general in the industry, kind of died down several years ago," adding that "it was hot for a minute and then it wasn't."

As aforementioned, audience perception and engagement were not popular topics of discussion among the nine media professionals. However, the following two interview excerpts—though unrelated—were considered important to address in the findings section. When Journalist 8 spoke of his Twitter bot that automatically tweets CanLII court judgements for his respective province, he also mentioned building another Twitter bot around the same time. This second bot, generated tweets on the fake weather reports shared by a well-known Quebec comedian. "It was a funny bot, and I made it clear in its Twitter bio that this account was purely designed for comedy." While comparing both Twitter bots—the comedy bot and the court ruling bot—Journalist 8 noticed that the former had many followers and the latter, barely any. Yet, he noticed that "over time, the comedy bot started losing followers while the other was gaining an important number of followers." He added that this experiment revealed that maybe, "automated processes do have value and do drive public engagement."

The final observation made by Journalist 1, had less to do with direct audience engagement and more to do with the dangers of inherent bias in the data that's being fed into news algorithms:

> Did you see that story about how self-driving cars are more likely to hit people of colour than white people because the data was bias? They tested the cars with white people because they only had white people working there, so their results were bias. That is a huge disservice to your audience […] I think the problem with these sorts of biases is that it's far reaching and it becomes a disservice to an audience—especially now that people don't read or consume news in the same way that they used to. They just grasp onto the easiest piece of information that they can retain and propagate it on Twitter and Facebook. So if you aren't verifying your data, if it isn't 100% crystal clear before it is used in a story, then you're disinforming.

Moreover, Journalist 1's observations circle back to important discussions—as presented in this findings chapter—that were had with other participants regarding the trustworthiness of data, algorithmic bias and accountability. These topics, which will be further explored in Chapter 5, serve this thesis toward the creation of ethical guidelines in relation to automated journalism. In addition, Journalist 1's concern about audience reception of news stories that are the product of faulty data or inherent biases further advance this thesis' emphasis on increased cross-media awareness among all actors involved in supporting the bottom-line of a media organization's newswork (Lewis & Westlund, 2015). While Journalist 1 claims that such biases become "a disservice to an audience," the next chapter reviews how it ultimately extends to being a disservice to media organizations, as they grapple with issues of authority and accountability in the age of algorithms and AI.

Chapter Five: Discussion and Conclusion

This chapter reiterates this thesis' objectives, which are complemented by a discussion of the findings with respect to best newsroom practice surrounding automated journalism, and journalist training and education in this sector. The findings are subsequently used to outline the possible development of ethical guidelines aimed at Canadian journalists and journalism educators. Finally, components of the research methodology and findings are further examined to put forward suggestions to future approaches to research in this field.

No known critical research has investigated who in Canadian media is implementing AI and newsroom automation, and the effects this has on the roles of Canadian journalists, their practices, methods, and tools. As such, the purpose of this thesis was to establish an original comprehensive overview of automated journalism and to determine where it fits within the country's news media landscape. Indeed, this thesis' exploration of journalistic production bolstered by automated newsroom technologies can be considered the first study dedicated to assessing the extent of automated journalism's Canadian adoption, thus adding to the discourse on the subject. The four questions guiding this thesis are the first, to the researcher's knowledge, to successfully address these gaps in the literature (see Introduction). In addition, the interviews with the nine news media professionals (see Table 1, Chapter 3; Chapter 4) successfully contextualize the current state of automated journalism in Canada by drawing inferences from the lived-experiences of these journalism actors with respect to developing, managing and applying various forms of this technology to their newswork (Kvale, 1996, p. 30).

Canada has been recognized worldwide as one of the global leaders in the advancement of AI research and development (see Introduction). Still, the lack of research positioning Canada within the automated journalism discourse led this thesis to correctly assume this field to still be

in its nascent stage within the Canadian news sector. Moreover, an analysis of published accounts of Canadian newsroom automation was not possible,[8] which is why interviews with working journalists and editors who are deeply familiar with automated journalism were necessary in order to advance the body of knowledge pertaining to this field. As discussed in Chapter 4, this thesis' findings resulted in several key aspects of research on automated journalism that fell in line with or diverged from the existing literature.

5.1. Summary of Major Findings

The major findings that coincide with previous studies are summarized as follows (see Chapters 1 and 4): 1) the human vs. machine cognitive intelligence debate is still prevalent today, leading some participants to reiterate that displacement or job loss to some degree is inevitable, although the goal is never to replace human journalists (Turing, 1950; Dreyfus, 1993; Carlson, 2015; Graefe, 2016); 2) the motivations driving the adoption of automated journalism in Canada echo what has already been said, namely the ability to speed up, scale up (Diakopoulos, 2019, p. 241), boost productivity, and all of this at very low marginal costs (Broussard, 2015; Stray, 2016); 3) the types of tools being used and stories being produced remain limited to topics that are rich in structured data (Graefe, 2016; Dörr, 2016; Stray, 2016); and 4) the findings indicate discrepancies in policies pertaining to transparency and accountability, which further mobilizes the urgent need to establish ethical guidelines (Broussard, 2016; Montal & Reich, 2017).

[8] As stated in the methodology section (Chapter 3), prior to participant recruitment, this thesis was informed by extensive online research through Google's search engine to explore topics specifically pertaining to Canadian newsroom adoption of automated journalism. This research phase revealed that published accounts of automated journalism in Canada were limited and thus, would not have offered additional insight about the present object of study.

The efforts put forward by this thesis also reveal findings that deviate from the literature, namely: 5) the Canadian journalism industry is widely viewed as being a troubled sector severely lacking in resources, making it difficult to hire experts with these specialist skills; 6) the newsrooms exercising more advanced methods of automated journalism are highly dependent on two factors, namely a) external funding, and b) newsrooms already employing staff who went into journalism with prior coding/programming knowledge; 7) some newsrooms view automated journalism as an opportunity to a) find new sources of revenue, b) help Canada's failing local news sector, and c) increase newsroom collaboration; and finally, 8) interview participants worry that Canadian newsrooms are too quick to adopt these technologies, and are concerned that a lack of cross-awareness and coordination at the organizational level will leave algorithms unquestioned and untested, thus reinforcing ethical concerns (Lewis & Westlund, 2015; O'Neil, 2016).

These results suggest that automated journalism will continue to make its way into Canadian journalism, though the pace at which it will be adopted and to what extent remains unclear, as newsrooms currently exercising automation are all doing so at varying degrees and at different stages of the news production pipeline. However, talks with participants reiterated a widespread skepticism of new technologies within the field of journalism at the individual (journalist) and organizational (media organization) level, which might further decelerate the shift towards newsroom automation in Canada (see Chapter 4). This could be especially true, considering the swift backlash Microsoft received this past June, after its AI-enhanced system confused two mixed-race singers of British pop group Little Mix in an article on one of the singer's ongoing experiences with racism (Vincent, 2020). As aforementioned in the findings chapter, Microsoft announced its new business strategy earlier in May—replacing employees

responsible for MSN's (Microsoft's search engine and news website) content optimization, aggregation and story curation with AI—thus giving its AI system the complete authority and responsibility of editorial oversight (Baker, 2020). Not only was the AI system unable to differentiate between images of the two women of colour—its algorithms selected a picture of the band member with the darkest skin tone to pair with the news article. Here, the algorithms associated a story about racism to being black, thereby selecting a picture of the darkest-looking singer of Little Mix despite there being no mention of her in the article. While Microsoft released a statement saying "this was not a result of algorithmic bias" (Vincent, 2020), research has shown algorithms to do just that, especially when it comes to racial discrimination and disproportionally marginalizing the poor (O'Neil, 2016; Noble, 2018).

This type of mistake is exactly why human oversight remains essential, even as research and development into automated systems continue to move forward. Notably, the case illustrated by Microsoft relates to several tensions that arose during the discussions with interview participants. For one, the work responsibilities of the former Microsoft employees—content optimization, aggregation and story curation—correspond to the types of routine, repetitive tasks that interview participants described as being "literally robotic functions […] that are boring and mind-numbing" (Journalist 5), "tedious" (Journalist 8) and "automatable" (Journalist 9). Still, certain participants in upper management positions also stressed that unlike Microsoft, their goal was never to replace staff, but rather to assist them in certain tasks while keeping up with rising demands to produce more content.

5.2. The Development of Possible Ethical Guidelines

In addition to the points mentioned above, this thesis' findings reveal concerns about the inherent biases built into algorithms, as well as the fact that interviewees underscored that Canadian newsrooms lack the qualified experts needed in order to hold their algorithms accountable when mistakes occur and for their decision-making patterns. Nevertheless, these findings provide an opportunity to adapt ethical guidelines and newsroom policies in order to consider automated journalism's place in the dynamic relationship between journalistic actors and technology (Lewis & Westlund, 2015), thus allowing journalists and journalism educators to reconsider their norms and practices accordingly.

The following proposed ethical guidelines were developed based on this thesis' major findings and thus, informed by the literature review (Chapter 1) and the lived-experiences of the nine interview participants (Chapter 4). This section merely serves as a guiding framework for news media professionals, covering the following topics of ethical newswork with in-depth analysis and justifiability: cross-media awareness, data handling and algorithmic accountability. As such, the present study argues that Canadian media organizations and journalism educators could benefit from such an ethics model, as it reinforces the unavoidable fact that automated journalism demands a re-evaluation of journalistic work practices and values.

5.2.1. Enhanced cross-media awareness and coordination

The findings presented in Chapter 4 and the guiding theoretical framework (Chapter 2) serve as the point of departure for this thesis' proposed ethical guidelines. The interdependence of all persons within the media sphere is essential to the proper framing and functioning of automated journalism, further bolstering best workplace practices. Automated journalism, as demonstrated by this thesis, thus works as a "boundary object," connecting the various actors

making up a media organization (Carlson, 2015). For journalistic actors with programming skills, it allows them to develop and hone algorithms and AI-enhanced systems capable of assisting in, or producing, data rich news narratives (Carlson, 2015). At the organizational level, automated journalism allows news outlets to revaluate their "business logics" (van Dalen, 2012) and provides them with an opportunity to expand coverage, uncover complex investigative stories (Broussard, 2015), and aim for higher profit margins and lower production costs (Graefe, 2016).

Yet, discussions with interview participants reveal that media organizations face these rising forms of interconnectedness with difficulty, as training and workflow changes demand the reassessment of the social, cultural and business structures of journalism organizations. This idea is reinforced by Marjoribanks (2000, p. 582, in van Dalen, 2012), stating that "social factors and the institutional contexts of media companies and markets, such as the work culture, position of journalism unions or the relations between owners and workers will shape the way this technology will be adapted." As such, automated journalism can only become successfully integrated if newsrooms increase coordination and "cross-media awareness" among all actors supporting the bottom-line of their newswork, while simultaneously adapting their journalistic norms and values (Lewis & Westlund, 2015). In turn, such changes to their work practice will further substantiate and potentially safeguard the roles of every news person involved in the integration and application of these automated technologies, which could perhaps mitigate journalists' underlying fears of displacement (Carlson, 2015).

Data handling

In this age of big data, the "algorithmic revolution in knowledge production" (Anderson, 2012), in journalism and in other fields, would not have been made possible without the explosion of available data (Diakopoulos, 2014). This "data deluge" (Wu, Tandoc Jr., & Salmon, 2019) has led newsrooms to face an unprecedented volume of available data ripe for collection, analysis and dissemination. Against this backdrop and the subsequent adoption of newsroom algorithms, media organizations have many serious ethical questions to consider about the data driving their automated processes, especially if they wish to stay true to journalistic values.

Data origin, reliability and veracity

Automated text generation using natural language generation, web scraping, and data mining enabled through machine learning, are all bound to raise ethical issues, as these systems and processes rely on the input of data derived from external sources. As such, newsrooms must establish a set of rules and vetting processes with regards to data origin, as the quality of the output (news story) is highly dependent on the veracity of the data inputs (Dörr & Hollnbuchner, 2017; WAN-IFRA, 2019). Regardless of whether automated journalism is used or not, the trustworthiness of any news organization is contingent on its ability to present factual, verifiable information to audiences. Therefore, the same standards must be applied to data, as data also raises questions of transparency and accountability.

During the discussions with interview participants, the majority maintained that the types of stories produced using automated methods primarily consisted of data-driven stories about sports, finance and government. Although the literature contends that data extracted from these types of authoritative sources is probably "safe for automatic crunching," media organizations should put measures in place to regularly go over such data to check its reliability and accuracy

before it is fed to an algorithm (Kent, 2015). This is especially important if data is obtained from less authoritative sources (i.e., non-governmental or non-academic sources).

In addition, data bias can be problematic as it can be present even before it is processed by an algorithm. Dörr & Hollnbuchner (2017) explain that incomplete or disorganized data can limit the "information value" of a dataset, whereby the collection of incomplete data can lead to further biases at the output level. The issue of missing data leading to bias was a point raised by Journalist 1, who provided the example of the story about self-driving cars, which published that self-driving cars were more likely to hit people of colour than white people—only because none of the people testing the cars were people of colour (see Chapter 4).

Legal considerations

Where data is extracted from also raises important copyright and privacy issues that media organizations must take into consideration. As mentioned by several interview participants, Canadian journalism has not yet established far-reaching data ethics around the process of downloading and collecting data. This remains somewhat of a grey area for journalism on a global scale, as WAN-IFRA (2019) indicated in their report that, even if some information is made publicly available online, that does not necessarily mean that it can be freely collected and subsequently published by a news outlet.

Further, publishing information that could lead to the violation of an organization's or individual's rights to privacy was also addressed during the interviews as being another important topic to consider. While source protection remains a traditional journalistic value, data used for automated journalism creates a tension between source protection and the recent push for data transparency (Dörr & Hollnbuchner, 2017). Therefore, it becomes difficult for media

organizations to maintain consistent data transparency methods as not all data being fed into algorithms contain information subject to privacy laws. Nevertheless, journalists and newsrooms alike should adopt disclosure guidelines that would take the different types of data being collected into consideration. As such, newsrooms and journalists should fully disclose their data sources and codebook when appropriate and within legal bounds, as this form of transparency may reinforce the trustworthiness of their newswork. If full disclosure of data cannot be made due to privacy and copyright laws, it should be thus indicated—granted the data can be legally used in the first place. As Journalist 8 put it, "you have to be transparent at all times, because audience trust comes from transparency," and data is as integral to full transparency as the subsequent automated technologies involved in the processing and analysis of said data.

Algorithmic accountability

As outlined in Chapter 1 (Section 1.4.) and during the interviews (Chapter 4), algorithmic accountability is fundamental to modern-day ethical news practices involving automated journalism. While many newsrooms are eager about using automation, as "the possibilities [with] this are really limitless right now" (Journalist 5), the significant error made by Microsoft's AI system should serve as an example to all media organizations. If newsrooms follow Microsoft's lead, where the error was simply communicated as a glitch in the "experimental feature of the automated system" (Vincent, 2020), the future of journalistic knowledge production and labour in this sector will most likely be stonewalled. Not only is it a cautionary tale of how not to remove editorial oversight, but the lack of accountability for the AI system's racially discriminatory error amply justifies audience distrust for newsroom automation. Carlson (2015) puts it best: "After all, the economic and authoritative underpinnings of automated journalism

rest on its acceptance outside the newsroom." Microsoft's AI-system error and the lack of well-established algorithmic accountability guidelines therefore expose the "entrenched cultural conflict between equating technological development with progress and deep distrust of machines as dehumanizing forces" (Carlson, 2015). Moreover, this thesis puts forward suggested algorithmic accountability guidelines informed by the extensive literature on the subject, and further substantiated by the discussions with interview participants.

Algorithmic bias

Algorithms have been widely discussed in the literature as operating with biases, as they learn how to classify, filter, analyze, curate and make decisions based on the set of rules and definitions humans build into them (Anderson, 2011; Diakopoulos, 2014; Diakopoulos, 2016; Young & Hermida, 2015; Noble, 2018; Broussard, 2018). This becomes especially precarious in a field like journalism. As mentioned in the findings, some participants continue to place objectivity at the centre of their professional self-perception (Deuze, 2005), whereby informing people in "ways that are nonpartisan […] is where we get to inform people the best we can" (Journalist 1). Indeed, van Dalen (2012) acknowledges that journalists have skills—flexibility, creativity, the ability to develop human relationships—that will never be appropriated by algorithms, and yet, "these are different than the ones journalists generally refer to when they describe the skills that define their profession," such as factuality and objectivity.

If the uptake of automated journalism continues, so should the acknowledgement that algorithms unescapably generate biased output. Still, this is not meant to "demonize algorithms, but to recognize that they operate with biases like the rest of [journalists]" (Diakopoulos, 2014). Moreover, certain actions leading to algorithmic accountability can be taken by newsrooms to

investigate its algorithms and further mitigate biases and subsequent reporting errors, as conveyed below.

Algorithm audits

Performing regular algorithm audits, a term found in the literature (O'Neil, 2016, p. 208; Broussard, 2018, p. 194; Diakopoulos, 2019, p. 207), was underscored by the majority of the participants as being one of the essential pillars of journalism ethics in the age of algorithms. An audit maintains algorithmic accountability in newsrooms (see Section 1.4), as it involves regularly sampling algorithms to monitor the input-output relationship in order to investigate and identify the level of influence, mistakes, or biases built into an algorithm (Diakopoulos, 2014). Yet, this remains a challenge in Canada, as the participants emphasized that very few people in Canadian newsrooms know how to monitor and test software and code between, during and after the input and output stages of algorithmic reporting. Therefore, newsroom algorithms should be met with carefully defined monitoring parameters before they are even considered by media organizations, although newsrooms already utilizing such technologies should adapt their work practices to make room for such protocols as well. Moreover, this thesis suggests the following steps for newsrooms currently implementing, or considering, algorithms for news production:

1) Algorithms (whether built in-house or purchased from an external provider) should undergo rigorous testing before being launched into action

Some participants emphasized that their newsrooms try "every singly edge case to make sure [their algorithm] works" (Journalist 4). This also included discussions of "unit tests, integration tests, regression tests [and] all the testing you can do" (Journalist 6). Despite these valid efforts, all participants who mentioned conducting tests on their algorithms explained that

this form of testing takes a lot of time, even indicating that it can take an upward of three months to ensure that the algorithm is properly functioning. Due to this lengthy testing process, one journalist, whose newsroom has a multiple test and fact-checking system in place for code, said that they do not conduct these tests on every algorithm involved in a story "because it would take an absurd amount of time" (Journalist 9), adding that such verification systems are reserved for larger stories. However, this thesis suggests that no matter how simple a task or story, the algorithms driving such tasks/stories should all undergo equal amounts of testing, as they all serve in journalistic production to some capacity, especially if they increasingly take on the role of communicator (Guzman, 2018). As such, rigorous testing, no matter the degree of complexity of an algorithm, allows journalists to uphold moral and journalistic standards of factuality, and enables newsrooms to prepare themselves for any libellous output that may occur, as they will have a better understanding of the decision-making patterns that led to the output results. Thus, newsrooms will maintain journalistic authority over the automated content they are producing, while exercising optimal algorithmic accountability standards (Dörr & Hollnbuchner, 2017).

2) Once an algorithm is up and running, regular and mandatory software/code testing should be conducted by the person overseeing the newsroom algorithm—and at least one secondary specialist

During the interviews, many participants spoke of building and developing their newsrooms' algorithms themselves, and thus being integral actors in regularly overseeing and managing these products. While this remains an essential aspect of algorithmic accountability, this thesis argues that newsrooms would greatly benefit from adding another regulatory step to their algorithms' oversight: a second set of eyes. Although this thesis' preliminary research on algorithmic accountability did not find any mention of this type of secondary inspection, many of

the participants—who build their own algorithms—recognized this form of accountability to be severely lacking in newsrooms. One journalist questioned the extent to which Canadian newsrooms are testing, monitoring and reverse-engineering their algorithms, and even expressed doubts, fearing no such verification systems to be in place within media organizations already carrying out automated journalism in their daily news production.

Moreover, even if journalists, much like this thesis' interview participants, have the programming skills needed to build these algorithms, the human-element remains a key driver of algorithmic distrust (Carlson, 2015). To reemphasize, Journalist 1 stated that: "that's the part where I'm the iffiest about using AI. It's because it's created by humans. It's not fail-proof." Thus, perhaps having a secondary person carrying out input-output tests on algorithms would allow newsrooms to identify some, if not all, of the algorithm's inadvertently programmed intentions and inner workings, and revise their code accordingly.

This suggested form of verification and testing is offered up here as a best-case scenario, that is, if a newsroom has the available manpower to perform these multi-person algorithm audits. However, two factors were discussed as limiting the widespread adoption of such auditory systems: 1) no Canada-specific (nor any other country-specific for that matter) journalistic self-regulation guidelines specifically designed for automated content exist (WAN-IFRA, 2019), thus leaving algorithmic accountability standards loose and underdeveloped, and 2) the lack of qualified specialists in Canadian newsrooms was discussed by participants as being a major obstacle to algorithmic accountability. Furthermore, if a newsroom lacks the necessary resources to hire more than one journalist with these specialist skills, this thesis suggests that newsrooms could perhaps outsource their code/software testing and monitoring to an external specialist (i.e., programmer specialist or academic). Still, this may be subject to future research,

as this study did not consider the possible legal and regulatory implications involved with sharing a newsrooms' proprietary information (in-house developed algorithm/product) with external actors. Nevertheless, the steps suggested above could serve as a guiding framework for Canadian newsrooms as they move forward with automated journalism in their workplace.

Editorial oversight

Every interview participant suggested that editorial oversight—that is, human input—remains essential and necessary despite some processes of automated journalism exercising the ability to operate under full autonomy. While the above section discussed the importance of consistent and rigorous monitoring of algorithms in between the stages of input and output, participants stressed that the same must be applied to automated processes during production and prior to publication. Once more, this step was removed when Microsoft's AI-system gained total control of the task of overseeing content optimization and story selection for MSN's news page (Baker, 2020; Vincent, 2020). As previously stated, the error made by the system could perhaps have been avoided were there human journalists still in place to provide an editorial override of its algorithmic authority. Further, such shifts in authority (from human to machine) extend beyond Microsoft and into other newsrooms. Dörr & Hollnbuchner (2017) reported that the *Associated Press* (AP), which has been fully automating the generation of its corporate earnings reports since 2014 (see Chapter 1), no longer monitors every earnings report generated, as they find it to be too time consuming—much like Journalist 9, whose newsroom omits testing algorithms for certain stories for the same reason. This seems ironic, as one of the main selling points for automated journalism adoption is its ability to produce content at speed and scale (Graefe, 2016).

If automated technologies are integrated to speed up and scale up a media organization's production of news, therefore saving journalists from doing hours of laborious work (Graefe, 2016; Diakopoulos, 2019), how does a newsroom maintain its authority—especially in instances where errors occur? Such questions of authority and accountability suggest that automated journalism comes with important trade-offs that newsrooms must carefully consider, even more so if they are adopting these technologies to increase their output with reduced staffing. Although Canadian newsrooms incorporating automated journalism into their news production are nowhere near the volume and scale of production seen at Microsoft and AP, these are still important questions that must be discussed at the organizational level. Thus, this thesis argues that while automated journalism is still in its early developmental phase in Canada, now is the best time for Canadian media organizations to broach these questions as they consider the place of automated journalism in their newswork.

Nevertheless, in accordance with what was shared by all nine interview participants, this thesis suggests that newsrooms must at all times maintain a human element to their automated news production, whether their system operates fully- or semi-autonomously. It is crucial that regular verification of the output is maintained before content reaches the public. If newsrooms like the ones employing this thesis' interview participants mandate this form of editorial oversight, they will be less likely to find themselves with errors. Thus, consistent and rigorous human input allows media organizations to once again maintain production authority over their automated systems, while mitigating errors and preserving journalistic standards of accuracy, reliability and trustworthiness toward their audiences.

Authorship disclosure and transparency

Much like Montal & Reich's (2017) study on newsrooms' authorship and transparency regulations with regards to its automated systems, this thesis found Canadian newsrooms to be operating inconsistently, with some participants indicating that their newsrooms do not disclose when an article is generated via automated methods (Chapter 4). In aligning with Montal & Reich (2017), the present research further emphasizes the need to establish a structured policy for bylining algorithm-generated news characterized by full disclosure of its author (including who developed the algorithm), its data origins, its methodology, and identifying which aspects of the story in question are automated.

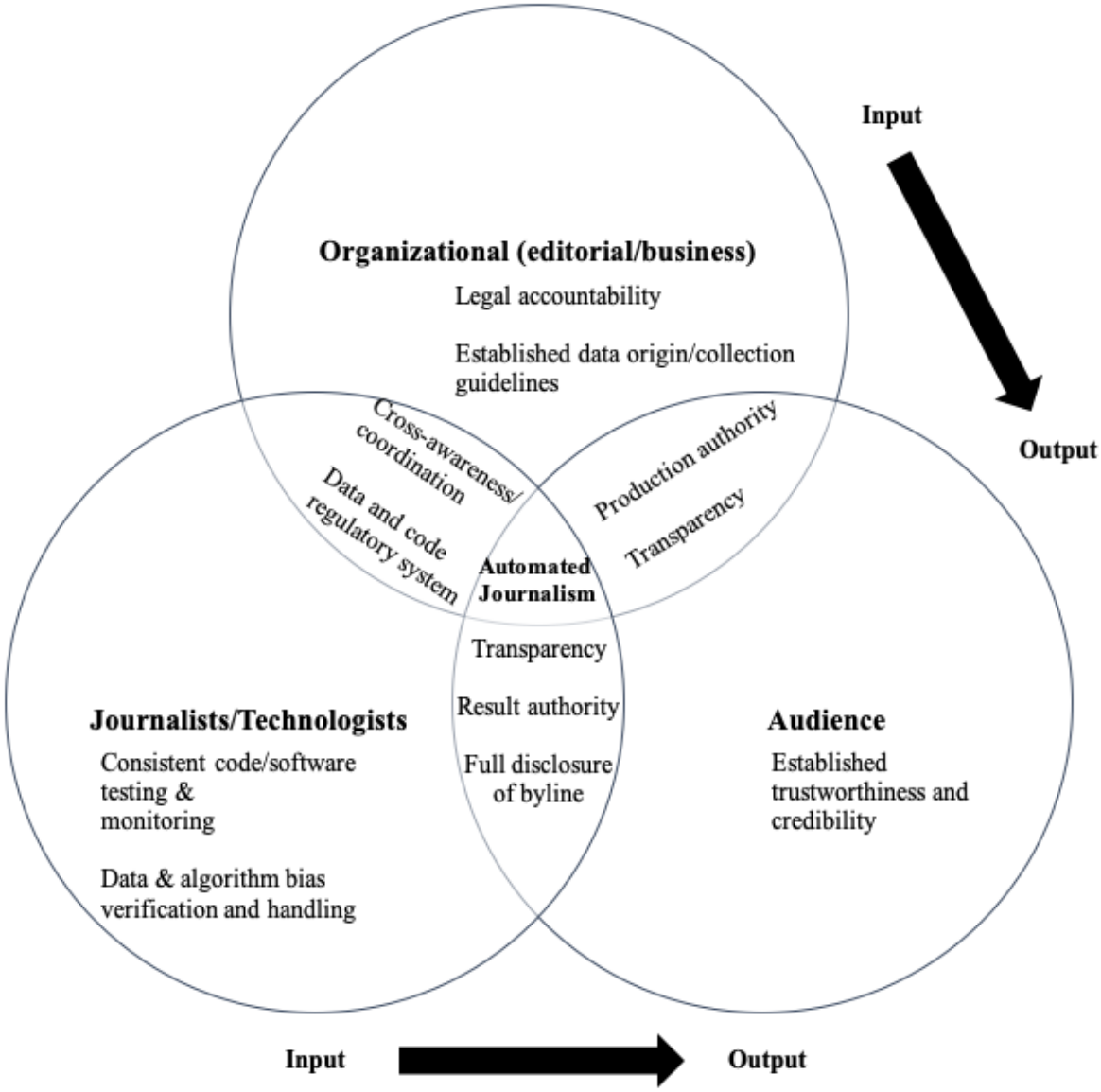
Thus, this form of transparency—fully disclosing all aspects of how a story was generated through automation—is particularly relevant to journalistic organizations as these technologies increasingly take on the roles of reporters. This form of transparency towards audiences, according to Montal & Reich (2017), is essential in that it can "mitigate the dangers of ambiguous authorship and demystify the naïve perceptions shared by audiences and journalists regarding the seemingly unbiased nature of algorithms." Along with Montal & Reich's (2017) comprehensive overview of bylining in the age of automation and algorithms, others have called on the increasing necessity to establish transparent disclosure policies (Diakopoulos, 2016; Broussard, 2016; O'Neil, 2016; Noble, 2018), although it remains unclear as to why newsrooms have not yet shifted their attribution policies. Thus, this thesis reemphasizes that such attribution policies reinforce transparency and accountability in view of "the public interest, journalistic values and practices," with the added value of reaffirming journalism's openness and trustworthiness toward audiences (Montal & Reich, 2017).

This form of transparency is also essential for news bots that directly interact with audiences on social media platforms, as news bots raise another important concern in relation to

the accountability of the bots themselves and "their creators and maintainers" (Lokot & Diakopoulos, 2016). While Journalist 9 mentioned during the interview that the "appetite for [ChatBots in the industry] kind of died down several years ago," the majority of interview participants provided examples of several types of news bots that they use in their regular reporting practices on Twitter and Slack (see Section 4.2.3., Chapter 4). Moreover, the study conducted by Lokot & Diakopoulos (2016) found that not all news bots in their study sample were transparent with regard to their sources, the algorithms behind their outputs, and even about the fact that audiences might indeed be interacting with bots and not humans. Indeed, Journalist 8 was adamant about the need for transparency about who is communicating with the public: "[newsrooms] need to be transparent about the fact that it's automated software engaging in communication with the public." Undoubtedly, audience perception and journalistic transparency are integral to the ongoing survival of knowledge production in the media sphere, which is why this thesis also reinforces the need for newsrooms' to fully disclose exactly who or *what* is transmitting information to its human audience (Guzman, 2018).

Furthermore, the ethical guidelines suggested above were informed by the in-depth discussions with interview participants and draw extensively from the literature. It is important to emphasize that the suggestions put forward are not exhaustive and merely serve as a starting point for media organizations, highlighting the benefit of such a holistic approach to automated newsroom practice. Nevertheless, this thesis provides an opportunity to adapt and prioritize newsroom guidelines by weighing the place of automated journalism within the news media landscape. This thesis' ethical framework is summarized on the following page (Figure 2).

Figure 2. Summary of proposed ethical guidelines for automated journalism



5.3. Strengths and Limitations

Limitations

Before presenting this thesis' final remarks, it is worth addressing the limitations of this project, which are subsequently followed by its strengths as well as suggested future directions. Firstly, Chapter 3 reiterates that the number of recruited participants fell short of the initial goal

of interviewing 12 to 15 news media professionals to fully saturate this thesis' collected data. While the response rate of nine recruited individuals was less than ideal, it is important to mention that all nine participants were male, although women known to be working with automated journalism methods in Canada were sent the initial thesis recruitment email as well. Moreover, this thesis was limited to the male bias of journalistic practitioners and could have benefitted from exploring the lived-experiences of women in the field. Journalist 10, who had to be excluded from the data and subsequent analysis given the fact that she was a data news developer working for an American news outlet with no experience in the Canadian sector, was the only working female journalist the researcher managed to speak to. She provided valuable insight on the state of automated journalism training and education in U.S. universities, thus revealing that this thesis and future research could benefit from focusing recruitment on obtaining more input from women in this sector. As such, further investigations into the state of automated journalism in Canada could explore the gender relationships pertaining to this field of study.

In addition, the recruitment of news media professionals only obtained individuals working in Quebec and Ontario, which proves to be a limiting factor as it does not provide insight into how automated journalism has made its way into media organizations in other parts of Canada, particularly the Maritimes and western provinces. While Journalist 4 provided concrete examples where he discussed developing automated newsroom tools for his former employer in Alberta (Chapter 4), this thesis remained limited to the lived-experiences of journalist practitioners in Quebec and Ontario, which suggests that future work is needed to address the scope of automated journalism on a larger scale, thus encompassing more of Canada's news media landscape.

Finally, while the goal of the present study was aimed at understanding the current and future implications of automated journalism on Canadian journalists and media organizations, there was little regard for how such technologies are perceived by Canadian audiences. As such, investigating audience reception of automated stories could further shed light on how audience members engage with these technologies, thus providing further knowledge and understanding of newsroom automation from a Canadian perspective.

Strengths

Although this study was restricted by the limited number of recruited participants, this thesis also exhibited strengths in its adoption of the hybrid thematic analysis method, which worked toward rigour and replicability by adopting both inductive and deductive approaches to thematic analysis (Fereday & Muir-Cochrane, 2006). As thoroughly discussed in Chapter 3, the lack of texts and resources on rigorous and relevant approaches to thematic analysis have led many researchers in recent years to attempt to demonstrate its value as being a method not tied to a particular epistemological or theoretical perspective, thus reinforcing its credibility as a method in its own right (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006; Maguire & Delahunt, 2017, Nowell, et al., 2017). As such, this thesis contends to having successfully fulfilled the qualitative requirements of data analysis by incorporating both Braun & Clarke's (2006) six-step framework to producing an effective thematic analysis and Fereday & Muir-Cochrane's (2006) dualistic method. Thus, this thesis was able to provide an insightful analysis that successfully answered this study's research questions (Chapter 1), and subsequently contributes to journalism scholarship on automated journalism, an area which has been lacking in knowledge contribution from the Canadian perspective.

As aforementioned, this study is the first of its kind, to the researcher's knowledge, to provide any sort of in-depth analysis on the current state of automated journalism in Canada. By combining Fereday & Muir-Cochrane's (2006) and Braun & Clarke's (2006) suggested approaches to conducting a thematic analysis, this thesis successfully contributes ethical guidelines for journalism practitioners and educators. It additionally highlights significant focus areas that should be considered for future attention, as will be discussed in the next section of this chapter. Finally, although the ethical guidelines (see Figure 2) on automated journalism presented here serve as a preliminary framework requiring further investigation, they suggest best directions for improved practices in the field, thus providing working journalists, educators and media organizations with guidance toward identifying and mitigating challenges that arise from the use of automated newsroom tools and methods. This is especially important as this thesis demonstrates that traditional ethical guidelines are no longer applicable to journalistic routines and practices, particularly if these technologies continue to make their way into the Canadian news media landscape.

Future directions

The use of a hybrid thematic analysis method may help advance future research in approaches to qualitative studies of journalism, and advance investigations of how automated journalism is produced and experienced by news media professionals, as well as audience members. While this thesis project's total number of interview participants was restricted by the limited knowledge of those applying automated journalism to their daily newswork in Canada, participants provided valuable insight on possible future directions for research in this field, namely on the implications of automated journalism at the local news level, as well as studies

exploring the adoption of these technologies to generate new avenues of revenue. These areas could benefit from further research and could expand on the preliminary ethical guidelines presented here.

In addition, while two of the interview participants also taught journalism courses at the university level, future research may consider exploring the extent to which journalism departments are carving out spaces for automated journalism in their curricula. This may be especially relevant as interviewees emphasized that today's journalism market is highly competitive and requires specialist skills beyond the basic elements taught in traditional journalism courses. Not only should future research look into the types of coding courses being offered, but perhaps the ways in which automated journalism shifts the journalistic skills and ethical values taught in school as universities set up students to become future reporters within this automated and technocentric news environment. Similarly, future research could explore continuing professional development, as interview participants highlighted this area to be especially challenging for media organizations as they reorient their business strategies towards newsroom automation.

5.4. Final Conclusion

This thesis project set out to investigate the current and future implications of automated journalism on the Canadian news sector, with a particular focus on the ways journalistic skills and values are shifting to make way for these increasingly used technologies. By reviewing the literature on the subject (Chapter 1), this study began by assessing the limited knowledge on the state of automated journalism in Canada, and developed four research questions aimed at filling this knowledge gap in the literature (see Introduction). Journalists and professionals holding

upper management positions within media organizations were recruited based on their deep understanding and application of these technologies in their daily news production (Chapter 3). Following the recruitment phase of this study, the participants went through in-depth semi-structured interviews in order to gain their interpretations of, and experience with, automated newsroom technologies (Chapter 4).

The data analysis was conducted using the hybrid thematic analysis method, which informed the proposed ethical guidelines for a more holistic approach to the application of automated journalism in the newsroom. In conclusion, this thesis' findings and subsequent proposed ethical guidelines should be considered a starting point for news media professionals currently implementing or considering the adoption of automated journalism. This thesis reiterates what has been said in past literature on the subject, in that it calls for a revaluation of journalistic norms and practices. However, it also provides a Canada-specific focus on automation, thereby contributing important findings about the country's nascent experience with automated journalism and elaborates future directions for news media professionals and educators if they wish to move forward with these technologies with what was said to be a "troubled industry" as the backdrop. Moreover, this study can help future research and current practitioners improve their newsroom practice, and thereby enhance the quality of Canadian automated journalism.

Finally, before embarking on this study, little was known about Canadian newsroom automation, but the interview participants offered immensely valuable insight that went far beyond the researcher's expectations. Thus, the author of this thesis hopes that her knowledge contribution about automated journalism serves as a catalyst for what will hopefully be a growing area of research within the field of journalism in Canada and elsewhere.

References

Adams, W. C. (2016). Conducting semi-structured interviews. In Newcomer, K. E., Hatry, H. P. & Wholey, J. S. (Eds.), *Handbook of Practical Program Evaluation,* (pp. 492-505). Jossey-Bass, San Francisco, California.

Akst, D. (2013). "What can we learn from past anxiety over automation?" *The Wilson Quarterly.*

Ananny, M. (2016). "Towards an Ethics of Algorithms." *Science, Technology & Human Values,* 41(1), pp. 93-117.

Anderson, C.W. (2011). "Deliberative, agnostic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency." *International Journal of Communication,* 5, pp. 529-547.

Anderson, C.W. (2012). "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society,* 15(7), pp. 1005-1021.

Anderson, C.W. (2013). "What Aggregators do: Towards a networked concept of journalistic expertise in the Digital Age." *Journalism,* 14(8), pp. 1008-1023.

Baker, G. (2020, May 29). Microsoft is cutting dozens of MSN workers and replacing them with artificial intelligence. The *Seattle Times*. Retrieved from https://www.seattletimes.com/business/local-business/microsoft-is-cutting-dozens-of-msn-news-production-workers-and-replacing-them-with-artificial-intelligence/

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks, CA: SAGE.

Braun, V. & Clarke, V. (2006). "Using thematic analysis in psychology." *Qualitative Research in Psychology,* 3, pp. 77-101.

Broussard, M. (2015). "Artificial Intelligence for Investigative Reporting." *Digital Journalism,* 3(6), pp. 814-831.

Broussard, M. (2016). "Big data in practice: Enabling computational journalism through code-sharing and reproducible research methods." *Digital Journalism*, 4(2), pp. 266-279.

Broussard, M. (2018). *Artificial Unintelligence: how computers misunderstand the world*. MIT Press, Cambridge, MA.

Browne, K. (2005). "Snowball Sampling: Using social networks to research Non-heterosexual women." *International Journal of Social Research Methodology: Theory and Practice,* 8(1), pp. 47-60.

Canadian Media Guild (GMC). (2013a). Jobs cut in the broadcast industry in Canada, November 2008 - August 2013. CWA/CSA. URL: https://www.cmg.ca/en/wp-content/uploads/2013/11/Preliminary-numbers-Broadcast-Job-cuts-between-2008-2013-CMG.pdf [May 25, 2020].

Canadian Media Guild (GMC). (2013b). Jobs cut in the print media industry in Canada, November 2008 - August 2013. CWA/CSA. URL: https://www.cmg.ca/en/wp-content/uploads/2013/11/Preliminary-numbers-Print-Media-Job-cuts-between-2008-2013-CMG.pdf [May 25, 2020].

Carlson, M. (2015). "The Robotic Reporter." *Digital Journalism,* 3(3), pp. 416-431.

Cohen, N. S. (2015). "From pink lips to pink slime: Transforming media labor in a digital age." *The Communication Review*, 2, pp. 98-122.

Crabtree, B. & Miller, W. (1999). A template approach to text analysis: Developing and using codebooks. In Crabtree, B. & Miller, W. (Eds.), *Doing qualitative research,* (pp. 163-177). Sage Publications Inc., Newbury Park, CA.

Creswell, J. W. & Poth, C. N. (2018). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches, fourth edition.* Sage Publications Inc., Thousand Oaks, CA.

DeCuir-Gunby, J. T., Marshall, P. L. & McCulloch, A. W. (2011). "Developing and using a codebook for the analysis of interview data: An example from a professional development research project." *Field Methods,* 23(2), pp. 136-155.

Deuze, M. (2005). "What is journalism? Professional identity and ideology of journalists reconsidered." *Journalism*, 6(4), pp. 442-464.

Diakopoulos, N. (2014). "Algorithmic Accountability Reporting: On the Investigation of Black Boxes." *Tow Center for Digital Journalism Brief, Columbia University.* Accessed on May 29, 2020. http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf

Diakopoulos, N. (2016). "Accountability in Algorithmic Decision Making." *Communications of the ACM,* 59(2), pp. 56-62.

Diakopoulos, N. (2018). "An algorithmic nose for news." https://www.cjr.org/tow_center/an-algorithmic-nose-for-news.php

Diakopoulos, N. (2019). *Automating the News: how algorithms are rewriting the media.* Harvard University Press, Cambridge, MA.

Di-Cicco, B. & Crabtree, B.F. (2006). "The qualitative research interview." *Medical Education,* 40, pp. 314-321.

Dörr, K. N. (2016). “Mapping the field of Algorithmic Journalism.” *Digital Journalism,* 4(6), pp. 700-722.

Dörr, K. N. & Hollnbuchner, K. (2017). “Ethical challenges of algorithmic journalism.” *Digital Journalism,* 5(4), pp. 404-419.

Dreyfus, H. L. (1993) *What Computers Still Can’t Do: A Critique of Artificial Reason*. MIT Press, Cambridge, MA. (first printed in 1972).

Eubanks, V. (2018). *Automating Inequality: how high-tech tools profile, police, and punish the poor.* St. Martin’s Press, New York, NY.

Fereday, J. & Muir-Cochrane, E. (2006). “Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development.” *International Journal of Qualitative Methods,* 5(1), pp. 80-92.

Glaser, B. G. & Strauss, A. L. (2009). *The Discovery of grounded theory: Strategies for qualitative research.* Transaction Publishers, Piscataway, NJ.

Graefe, A. (2016). *Guide to Automated Journalism.* Tow Center for Digital Journalism.

Graefe, A., Haim, M., Haarmann, B., Brosius, H-B. (2018). “Readers’ perception of computer-generated news: Credibility, expertise, and readability.” *Journalism,* 19(5), pp. 595-610.

Guzman, A.L. (2018). What is human-machine communication, anyway? In. A.L. Guzman (Ed.)., *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1-28). Peter Lang, New York, NY.

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea.* MIT Press, Cambridge, MA. (first printed in 1985).

Kent, T. (2015). “An ethical checklist for robot journalism.” https://medium.com/@tjrkent/an-ethical-checklist-for-robot-journalism-1f41dcbd7be2

Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing.* Sage Publications, Thousand Oaks, California.

Lewis, S.C. & Westlund, O. (2014). “Big data and journalism: epistemology, expertise, economics, and ethics.” 3(3), pp. 447-466.

Lewis, S.C. & Westlund, O. (2015). “Actors, actants, audiences, and activities in cross-media news work.” *Digital Journalism*, 3(1), pp. 19-37.

Lewis, S.C. & Westlund, O. (2016). Mapping the human-machine divide in journalism. In Witschge, T., Anderson, C., Domingo, D. & Hermida, A. (Eds.), *The SAGE Handbook of Digital Journalism* (pp. 341-353). SAGE Publications, City Road, London.

Lewis, S. C., Guzman, A. L. & Schmidt, T. R. (2019). "Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News." *Digital Journalism,* 7(4), pp. 409-427.

Linden, C-G. (2017). "Decades of automation in the newsroom: Why are there still so many jobs in journalism?" *Digital Journalism,* 5(2), pp. 123-140.

Lokot, T. & Diakopoulos, N. (2016). "News Bots." *Digital Journalism,* 4(6), pp. 682-699.

Maguire, M. & Delahunt, B. (2017). "Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars." *All Ireland Journal of Teaching and Learning in Higher Education,* 8(3), pp. 3351-33514.

Marconi, F., Siegman, A., & Machine Journalist. (2017). *The future of journalism: A guide for newsrooms in the age of smart machines.* New York: Associated Press. https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf.

Mihas, P. & Odum Institute. (2019). *Learn to build a codebook for a generic qualitative study.* SAGE Publications, London, United Kingdom.

Montal, T. & Reich, Z. (2017). "I, Robot. You, Journalist." *Digital Journalism,* 5(7), pp. 829-849.

Noble, S. (2018). *Algorithms of Oppression: How search engines reinforce racism.* NYU Press, New York, NY.

Nowell, L. S., Norris, J. M., White, D. E. & Moules, N. J. (2017). "Thematic analysis: Striving to meet the trustworthiness criteria." *International Journal of Qualitative Methods,* 16, p. 1-13.

O'Neil, C. (2016). *Weapons of Math Destruction: how big data increases inequality and threatens democracy.* Broadway Books, New York, NY.

Pavlik, J. (2000). "The impact of technology on journalism." *Journalism Studies,* 1(2), pp. 229-237.

Rice, P. & Ezzy, D. (1999). *Qualitative research methods: A health focus.* Melbourne: Oxford University Press.

Roberts, K., Dowell, A. & Nie, J-B. (2019). "Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development." *BMC Medical Research Methodology,* 19(66), pp. 1-8.

Ryan, G. & Bernard, H. (2000). Data management and analysis methods. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 769-802). SAGE, Thousand Oaks, CA.

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE, Thousand Oaks, CA.

Signorelli, C. M. (2018). "Can Computers Become Conscious and Overcome Humans?" *Frontiers in Robotics and AI,* (5), pp. 1-20.

Smith, A. & Anderson, J. (2014). *AI, Robotics, and the Future of Jobs.* Pew Research Center

Stray, J. (2016). "The age of the cyborg." *Columbia Journalism Review,* Fall/Winter 2016, https://www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php

Thurman, N., Schiffers, S., Fletcher, R., et al. (2016). "Giving Computers a Nose for News." *Digital Journalism,* 4(7), pp. 838-848.

Turing, A. M. (1950). "Computing Machinery and Intelligence." *Mind* 59(236), pp. 433-460.

United States Bureau of Labor Statistics (2018). *Occupational employment statistics.* URL: https://www.bls.gov/oes.tables.htm [May 25, 2020]

Usher, N. (2016). *Interactive Journalism: hackers, data, and code.* Urbana: University of Illinois Press.

Van Dalen, A. (2012). "The algorithms behind the headlines: how machine-written news redefines the core skills of Human Journalists." *Journalism Practice,* 6(5-6), pp. 648-658.

Vincent, J. (2020, June 9). Microsoft's AI journalists confuse mixed-race Little Mix singers on MSN homepage. *The Verge.* Retrieved from: https://www.theverge.com/2020/6/9/21284934/microsoft-ai-news-editors-msn-homepage-little-mix-singers

Wilkinson, S. & Winseck, D. (2019). "Crisis of transformation? Debates over journalistic work in Canada." *Canadian Journal of Communication,* 44(3), pp. 373-395.

Wölker, A. & Powell, T.E. (2018). "Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism." *Journalism,* pp. 1-18.

Woolley, S., Boyd, D. Broussard, M. (2016). "How to think about bots," https://motherboard.vice.com/en_us/article/qkzpdm/how-to-think-about-bots

World Association of Newspapers and News Publishers. (2019). *News Automation: The rewards, risks and realities of 'machine journalism.'* Frankurt, Germany: WAN-IFRA.

http://immersiveautomation.com/wp-content/uploads/2019/06/WAN-IFRA_News_Automation-FINAL.pdf

Wu, S., Tandoc Jr., E. & Salmon, C. T. (2019). "Journalism reconfigured." *Journalism Studies,* 20(10), pp. 1440-1457.

Young, M. L. & Hermida, A. (2015). "From Mr. and Mrs. outlier to central tendencies." *Digital Journalism,* 3(3), pp. 381-397.

Zamith, R. & Braun, J.A. (2019). "Technology and Journalism." *The International Encyclopedia of Journalism Studies*

Appendices

Appendix I. Semi-structured interview script

Description and consent:

- This research project generally looks at the current practices and future implications for automated journalism and algorithmically-driven reporting in Canada. It seeks to critically and systematically account for automated journalism's place in the Canadian news media landscape and examines the possibilities and challenges algorithmically-driven reporting practices present to Canadian journalists and other news media professionals. I felt it was important to speak to journalists who actually use automated journalism methods in their work or newsroom, to get a better understanding of how automation impacts journalistic skills, processes and stories. So, I have a few questions I'd like to go over about how you do your work, how you integrate automated journalism in it, and the journalism skills you feel are important as automated journalism becomes increasingly integrated.

- But before going ahead, I'd like to make sure you understand your participation in this project. It's outlined in the consent form, which I'll ask you to sign before we get started. Have you had a chance to read the form?

(Give time if not)

Do you have any questions about the interview, or my project as a whole?

- I want to make sure you understand that your participation is completely voluntary, and that you can withdraw at any point without any consequences. If you choose to withdraw from the study, please let me know at any time within **one month** of your interview date in order to withdraw your participant data. Should you opt out, all data will be withdrawn from analysis and destroyed. I will also take steps to keep participant identities confidential in all of my research reports. I won't be including any identifying information outside of, for example "a newsroom journalist working for a Montreal-area news organization with X number of years' experience in the field." So, anything you say here won't be directly linked back to your name. Also, my faculty supervisor and I will be the only ones who will have access to the audio recordings and the transcriptions thereof. Finally, if there are any of my questions you don't have an answer to or feel uncomfortable answering, that's fine. In other words, you don't have to answer if you don't want to.

(Sign consent forms)

(Start recording if okay)

Introductions:

- Can you please state your name, your job title, and how long you have been working as a journalist?

- Can you please state today's date, and where you are?

Questions:

1. What does being a journalist mean to you?

2. In your opinion, are journalists generally receptive of new technologies in the newsroom? Why or why not?

3. How would you define automated and semi-automated journalism in the context of your work?

4. What types of algorithmically driven or automated tools and processes do you use in your newsroom?

5. When did you begin adopting these types of journalism tools and what were your motives behind adopting them?

6. What types of stories or tasks does automated journalism help you tell or accomplish?

7. How has the application of automated journalism in your newsroom changed your job?

8. How has automated journalism, if at all, shifted journalistic labour and skills?

9. Do you view automated journalism as a benefit or threat to your profession? More specifically, a threat to those who don't practice it?

10. How, if at all, has the use of automation and algorithmically driven tools changed authorship and transparency guidelines?

11. Do you think newsrooms should adopt policies/strategies surrounding these technologies? Why/why not?

12. In your opinion, do we need to establish data journalism ethics around the process of scraping or downloading data from non-government websites? Why/why not?

13. How are the algorithms driving the content and stories produced held accountable?

14. In your opinion, what are the advantages and disadvantages of using these tools to communicate and interact with your audiences?

15. Based on your knowledge, do you think Canada is behind in adopting automated journalism? Why/why not?

16. Do you think the existing language barrier between Quebec and the ROC makes it difficult to train people in this field of journalism?